Demand Research, LLC

Macro Consumption Metrics Pilot Study
Final Report

by

Marvin J. Horowitz, Ph.D.

Prepared for:

California Public Utilities Commission
Energy Division

November 8, 2012

Table of Contents

EXECUTIVE SUMMARY

This report describes the findings of the *Demand Research* Macro Consumption Metric Pilot Study. These findings focus on the reductions in energy use attributable to the energy efficiency policy (the combined collection of energy efficiency programs, building codes, appliance standards, and other public initiatives) in California in between 2006 and 2010. The data used for this study encompass over 6,000 California census tracts that make up the service territories of PG&E, SDG&E, and SCE. The source of the energy consumption data are the IOU's monthly customer billing data that are annualized, address-normalized, and merged by 2010 census tracts. The data span 2006 through 2010.

Using econometric models designed with the same basic structure, policy impacts are estimated for the PG&E and SDG&E residential sector both for electricity and natural gas consumption (residential sector data for SCE and SCG were not available for this study but will be included in the study database for use in future analyses). For the commercial and industrial sectors, electricity policy impacts are estimated for PG&E, SDG&E, and SCE collectively at the county level. Natural gas policy impacts for these sectors are estimated for PG&E and SDG&E. Also, the impacts of residential building codes on housing units built between 2000 and 2004 are estimated both for electricity and natural gas consumption.

The upper panel of Table ES1 contains the electricity efficiency policy impact findings for the four year period from 2006-2009; the lower panel contains the findings for the five year period from 2006-2010. Table ES2 is in the identical format and displays the findings for natural gas efficiency policy.

Based on the collected findings of the eight electricity consumption econometric models estimated for this study, in 2009 the total cumulative impact of electricity efficiency policy in all sectors between 2006 and 2009, including residential building code impacts for housing units built between 2000 and 2004, is a reduction in total electricity use of 8,355 GWh. This is a 5.4 percent decline relative to the average, total energy consumption per year in the 2006-2009 period. The relative standard error of the impact estimate is 28.6 percent; this translates to a 90 percent confidence interval of between 4,489 and 12,282 GWh of electricity savings. By comparison with the econometrically-derived savings estimate, in 2009 the cumulative ex ante estimate of electricity reductions for the 2006-2009 period due to downstream, IOU-implemented energy efficiency programs is 3.9 percent. It is not the purpose of this study to speculate as to why there are differences in the ex ante and econometric model-based energy consumption reduction estimates. Suffice it to say that the cumulative ex ante estimates of energy reductions due to downstream, IOU-implemented energy efficiency programs are used in this study as *indicators* of the impacts of the broader set of public initiatives that comprise *de facto* state-wide energy efficiency *policy*.

Table ES1: State-Level Policy Impact Findings, Electricity

kWh: 2006-2009 Period	Avg. Ann. Energy Consum.	Last Yr. Cum. Ex Ante Savings	Pct. Est. Impact	Annual Est. Impact
Ind. (PG&E, SDG&E, SCE)	41,554,138,628	2,048,456,628	13.6%	5,668,443,641
Com. (PG&E, SGG&E, SCE)	76,735,020,087	3,403,092,575	1.2%	916,539,984
Res. (PG&E)	30,132,043,300	480,430,254	4.7%	1,408,088,335
Res. (SDG&E)	7,483,267,512	105,504,485	3.9%	289,567,961
Res. Codes (PG&E)[1]	2,571,722,287		1.9%	48,775,285
Res. Code (SDG&E)[1]	742,521,437		3.2%	23,700,542
Total	155,904,469,527	6,037,483,943		8,355,115,748
Percent Impact		3.9%		5.4%
Pct. Standard Error (+/-)				28.6%
kWh: 2006-2010 Period				
Ind. (PG&E, SDG&E, SCE)	41,879,937,508	2,512,486,691	17.8%	7,471,169,145
Com. (PG&E, SGG&E, SCE)	76,829,480,556	4,412,722,613	2.1%	1,611,926,492
Res. (PG&E)	30,207,548,725	526,324,700	6.4%	1,923,104,970
Res. (SDG&E)	7,432,945,519	129,427,756	4.2%	313,568,580
Res. Codes (PG&E)[1]	2,552,293,975		2.0%	51,114,791
Res. Code (SDG&E)[1]	729,653,042		2.9%	21,086,973
Avg. Total. Ann. Consum.	156,349,912,308	7,580,961,761		11,391,970,952
Percent Impact		4.8%		7.3%
Pct. Standard Error (+/-)				18.9%

[1] Average annual energy consumption for census tracts with >30% houses built 2000-2004 .

Cumulative policy impacts for the 2006-2010 period are 7.3 percent. The relative standard error of this estimate is plus or minus 18.9 percent, or 31 percent at the 90 percent confidence level. The cumulative IOU energy efficiency program ex ante energy reduction estimate is 4.8 percent of average total energy consumption over the five year period.

Table ES2 indicates that the eight natural gas consumption econometric models yield findings of 1.9 percent policy impacts over the four year estimation period.

Table ES2: State-Level Policy Impact Findings, Natural Gas

Therms: 2006-2009 Period	Avg. Ann. Energy Consum.	Last Yr. Cum. Ex Ante Savings	Pct. Est. Impact	Annual Est. Impact
Ind. (PG&E,SDG&E)	5,124,383,847	87,353,504	1.1%	56,171,929
Com. (PG&E, SDG&E)	1,198,857,181	17,121,520	18.6%	222,477,711
Res. (PG&E)	2,012,166,142	10,406,303	1.2%	23,976,294
Res. (SDG&E)	297,293,772	2,039,215	-51.3%	-152,541,395
Res. Codes (PG&E)[1]	152,939,950		9.2%	13,873,030
Res. Code (SDG&E)[1]	25,880,855		5.1%	1,328,257
Total	8,632,700,942	116,920,542		165,285,826
Percent Impact		1.4%		1.9%
Pct. Standard Error (+/-)				175%
Therms: 2006-2010 Period				
Ind. (PG&E,SDG&E)	5,143,530,663	111,485,401	1.4%	72,730,412
Com. (PG&E, SDG&E)	1,200,231,263	23,026,253	19.5%	234,511,830
Res. (PG&E)	2,018,224,763	13,930,364	3.3%	65,597,381
Res. (SDG&E)	308,983,896	2,778,731	-64.3%	-198,590,214
Res. Codes (PG&E)[1]	152,601,314		9.6%	14,650,642
Res. Code (SDG&E)[1]	25,675,625		4.1%	1,045,537
Total	8,670,970,586	151,220,748		189,945,588
Percent Impact		1.7%		2.2%
Pct. Standard Error (+/-)				244%

[1] Average annual energy consumption for census tracts with >30% houses built 2000-2004 .

The relative standard error of this estimate is 175 percent. The IOU energy efficiency program ex ante estimate of natural gas reductions for the four year period is 1.4 percent relative to average total electricity consumption. In the five year period, the impact estimate for natural gas consumption energy efficiency policy rises to 2.2 percent, again with a large relative standard error.

This study achieves the two main goals of this pilot study articulated by Commission Decision (D.)10.10.33 (October 28, 2010). Both are related to the creation of an evaluation framework that is scientifically defensible and applicable for the foreseeable future. First, it demonstrates that a well-founded econometric framework, coupled with an appropriate, large-sample database, can be developed to evaluate the aggregate impact of the 2006-2008 energy efficiency programs on energy consumption. Second, it demonstrates that aggregate econometric models employing large samples are capable of accurately measuring the impact of the Commission's energy efficiency efforts on overall electricity and natural gas consumption in California in the context of post-2012 EM&V activities. The potential for accurate measurement is demonstrated by the standard errors that accompany the estimated electricity policy impacts for the 2006-2010 period. No other type of evaluation study can produce a relative error bound of 31 percent (at the 90 percent confidence level) around a state-level policy impact estimate that embraces all three non-transportation sectors of the economy and incorporates the uncertainties due to free ridership, spillover, rebound, measure interaction and retention, behavioral changes, and general economic conditions.

This study also supports discussion of the two additional goals articulated by Decision (D.) 10.10.33. This detailed, small-geographic area, sector and industry-level approach to policy evaluation shows that such studies are likely to be valuable for improving estimates of aggregate reductions in Greenhouse Gases (GHG) emissions from efficiency programs as required in AB32. Also, it is likely that they can prove valuable for more directly aligning and integrating energy efficiency program findings into the California Energy Commission's (CEC) demand forecasts, and ultimately, the CPUC's resource procurement process. In any event, detailed discussions of these goals are not the purpose of this pilot study; the Commission has contract with another party to examine these goals.

General recommendations for integrating this evaluation approach into the permanent portfolio of post-2012 EM&V activities fall into two categories, database development and econometric analysis. They include:

- Expand the database with additional variables and upgrade the database for easier access.
- Develop standardized routines for data cleaning and checking.
- Develop and evaluation-oriented geographic information system.
- Explore the properties of different types of econometric impact estimators.
- Experiment with customized models for different fuels, sectors, utility service territories, market segments and customer grouping.
- Develop econometric models that target specific programs and public initiatives.
- Experiment with analyzing census tract-level end use load shapes derived from smart meter equipment.

1. Introduction

This report describes the findings of the *Demand Research* Macro Consumption Metric Pilot Study. These findings focus on the reductions in energy use attributable to the energy efficiency policy (the combined collection of energy efficiency programs, building codes, appliance standards, and other public initiatives) in California in between 2006 and 2010. The data used for this study encompass over 6,000 California census tracts that make up the service territories of PG&E, SDG&E, and SCE. The source of the energy consumption data are the IOU's monthly customer billing data that are annualized, address-normalized, and merged by 2010 census tracts. The data span 2006 through 2010.

Using econometric models designed with the same basic structure, policy impacts are estimated for the PG&E and SDG&E residential sector both for electricity and natural gas consumption (residential sector data for SCE and SCG were not available for this study but will be included in the study database for use in future analyses). For the commercial and industrial sectors, electricity policy impacts are estimated for PG&E, SDG&E, and SCE collectively at the county level. Natural gas policy impacts for these sectors are estimated for PG&E and SDG&E. Also, the impacts of residential building codes on housing units built between 2000 and 2004 are estimated both for electricity and natural gas consumption. As articulated in Commission Decision (D.) 10.10.33 (October 28, 2010), there are five primary goals to the overall Macro Consumption Metrics project:

1) To assess the ability of total energy consumption approaches to accurately measure the aggregate impact of the 2006-2008 energy efficiency programs on energy consumption.
2) To assess the ability of total energy consumption approaches to accurately measure the impact of the Commission's energy efficiency efforts on the overall electric energy and natural gas consumption in California in the context of post-2012 EM&V activities.
3) To examine the ability of total energy consumption approaches to improve estimates of aggregate reductions in Greenhouse Gases (GHG) emissions from efficiency programs as required in AB32.
4) To examine the ability of total energy consumption approaches to more directly align and integrate the study results into the California Energy Commission's (CEC) demand forecasts and ultimately the CPUC's resource procurement process.

5) To provide recommendations as to the specific data needs, analytical frameworks, and systems required to integrate total energy consumption approaches into the permanent portfolio of post-2012 EM&V activities.

Demand Research's pilot study explicitly addresses the Commission's first goal and thereby implicitly addresses the second one, the value in carrying this evaluation approach into the future. Assuming no major changes in the direction of the Commission's energy efficiency program efforts, the findings of this study speak for themselves regarding the value in continued development and refinement of the project database and methodology. This leads addressing the fifth Commission goal by providing, at the end of this report, a list of specific improvements in data and modeling that would help to integrate this evaluation approach into the permanent portfolio of post-2012 EM&V activities. The Commission contacted with a party other than *Demand Research* to address goals three and four of the Macro Consumption Metric project.

In many respects, the present modeling effort is similar to past econometric studies of aggregate energy consumption, most of which have been cited in three independent white papers on this subject produced in 2011 for the Commission. Like most prior studies, the econometric models used to analyze aggregate energy use are populated with cross section, location-specific observations whose variables are measured at two or more equal time intervals. Yet, the present study introduces many new research design features with the potential for greater development. These begin with an innovative approach to inexpensive data collection. The key features of the database created for this study are:

- Census tract-level aggregated electricity and natural gas consumption data for the five years from 2006 to 2010. In addition to census tract level electricity and natural gas consumption data, downstream (end user) IOU-implemented energy efficiency program data are available for key variables such as ex ante energy reductions, total measure costs, and IOU incentive costs per measure.
- Commercial and industrial sector energy consumption data disaggregated into NAICS-based industry categories at the county level.
- Annual small-area climate data, population and housing data for each census tract, and county, state, and national economic data.

The richness of this database permits certain types of statistical analyses to be performed for the very first time. These demonstrate that this evaluation methodology can enhance future energy efficiency policy development and evaluation efforts, energy forecasting and resource planning efforts, and environmental monitoring efforts. The following are definitions of key concepts and terminology used in this pilot study:

a) *Energy consumption*: Electricity and natural gas consumption are represented in the estimated models by utility billing data and does not include self-generation. The expression "energy use" is used synonymously with energy consumption.

b) *Ex-ante reductions in energy consumption:* The gross energy reductions reported in IOU energy efficiency program databases that are assumed to be realized from energy efficiency measures installed via IOU-run energy efficiency programs.

c) Energy efficiency policy: Policy is an umbrella term that refers to the full collection of sector-specific energy efficiency programs and public initiative that operate simultaneously in a given location. These initiative may or may not be coordinated with each other. Energy efficiency building codes and standards are one element of energy efficiency policy.

d) *Cross section, time series,* and *panel studies:* Cross section studies are made up of subjects, such as households, companies, or groups of subjects, for whom data are collected for one or two time periods. Time series or longitudinal studies are made up of data collected for multiple, equally-spaced time periods for a single subject or single group of subjects. Panel or pooled studies combine the two. They are made up of multiple subjects or multiple groups of subjects for whom data are collected for multiple time periods. The advantage of a panel study is that by combining information on how energy consumption changes from year to year (the time series component), with information on how energy consumption differs from subject to subject (the cross section component) it offers more comprehensive insights into long-term changes in energy consumption than any other type of study.

e) *Policy Impacts*: In the context of this study, policy impacts are econometric-based estimates of the reductions in consumption attributed to energy efficiency policy. Although statistics derived from IOU program tracking systems of downstream programs are used in the econometric models, they are interpreted broadly as indicators of policy impacts, not as indicators of the impacts of downstream programs alone. Since the econometric models of energy consumption control for market factors such as incomes, prices, and weather, and since the models are estimated over four or five years periods, the policy impacts are interpreted as long-term energy consumption changes that exclude free ridership and include spillover, rebound, measure interactions, or other externalities. As such, they may be considered *net* savings in the truest sense of the word.

A concept that is not broached in this study is that of *total market gross energy savings.* This concept, used by the CPUC for program planning and goal setting, is defined as the sum of projected naturally-occurring efficiency plus the sum of savings from all programs targeted to a specific population. There is no analog to this concept in the econometric models estimated in this study. By their nature, econometric models use historical data to provide estimates of how one or more variables influenced energy consumption.

Section 2 of this study describes the econometric modeling framework and the construction of relevant variables, Section 3 describes the policy impact findings for each sector and fuel, and Section 4 concludes with general recommendations for continuing to collect data, develop analytical frameworks, and develop the kind of system required to integrate total energy consumption approaches into the permanent portfolio of post-2012 EM&V activities.

2. Policy Modeling Framework

As this pilot study explores the use of aggregate energy consumption data to evaluate sector-level, fuel-specific, energy efficiency policy impacts on an annual, ongoing basis, the strategy of this study is to demonstrate the capability, and value, of developing a basic econometric modeling and research design framework. This approach differs from many econometric studies that are geared towards analyzing narrow technical issues using specialized tools that cannot be universally applied. The mission here is to develop a policy impact measurement approach that is relatively constant from year to year and subject to subject, thereby allowing policymakers and resource planners to be continuously informed of program accomplishments.

Toward this end, several major principles are followed throughout this study. For one, the analyses of electricity use and natural gas use are, for the most part, treated identically. This principle has its pros and cons. On the one hand it demonstrates the practicality and the validity of the approach, but on the other it sacrifices precision. For example, the same scheme that is used for the electricity consumption analysis for combining 24 industries into 13 industry categories is also used for the natural gas consumption analysis. Although different industry categories might lead to more accurate findings for both fuels, it might also lead to findings that are tied to the peculiarities of the data and will change over time. Therefore, to best assess the future potential of this evaluation approach, standardization is imposed whenever possible.

A second, related principle is that the impact estimator and basic model specifications be similar across the three economic sectors. The word *similar* is used purposefully, because it is impossible, not to mention unwise, to apply the same models to all sectors. Different variables are available for different sectors, and different variables drive the energy use and policy impacts in different sectors. Moreover, unexpected data issues arise, such as sector-level differences in the availability and accuracy of IOU-run energy efficiency program data. Thus, while the basic framework for modeling and analysis can be similar, the details necessarily vary.

The third and last major principle followed in this study follows from the two above. In plain language, it is to not lose site of the forest for the trees. At present, 16 models are estimated for this study for the purposes of learning more about what this new evaluation method, and these data, are capable of offering. Judging the merits of an individual model based on a single statistic or diagnostic test is beside the point, as is exploring why the coefficient of a variable such as "years of schooling" might have the expected sign and be statistically significant in three residential sector models with identical specifications, but not a fourth. Fine-tuning a single model is always possible. What is more important, and far more difficult, is to create an evaluation framework that is scientifically defensible and will be broadly applicable for the foreseeable future.

The following sub-sections describe the kind of model that is propagated throughout this study, the theory behind it, how the major variables in the models are constructed, and how policy impacts are estimated.

2.1 Panel Fixed Effects Models

All of the models estimated for this study are panel, fixed effects models. Panel models are those in which for each individual cross section unit there are data for two or more time periods. The time periods in this study are measured in years, and the maximum number or years of data that are available are five, from 2006 to 2010. The cross sections differ based on the sector that is being modeled. In the residential sector the cross section units are census tracts, and in the commercial and industrial sectors the cross section units are industries by county and by IOU.

Fixed time effects are implemented in panel models as dummy variables that differentiate each year from every other year, and fixed cross section effects are implemented as dummy variables that differentiate each cross section from every other cross section. These variables produce coefficients that are model intercept shifters; that is, they change the values of the coefficients that reflect all of the unobserved but systematic factors that affect the dependent variable. In particular, fixed time effects coefficients reflect idiosyncratic factors that are specific to a particular year but affect all cross sections. Conversely, fixed cross section effects, are specific to each cross section but not specific to any one year. Because fixed effects gather up the influences of all the variables that are unobserved (and are thus left out of the model) they are general corrections for model misspecification due to omitted variable bias. This is important to note because variables related to electricity and natural gas prices are not sufficiently disaggregated at the census and county levels to be incorporated into the models.

Besides the fixed effects variables, each panel model contains a number of continuous variables that are typically considered determinants of energy use, such as weather and income. The dependent variable in all of these models is *energy use per site*. It is important to note that utility customers can be counted in one of two ways; either by the officially listed *premises* being served, or by customer *accounts*. As publically-available independent variables are geared towards explaining the energy use per building or per location, the former is chosen to represent customers. It more accurately reflects the number of unique buildings being served by a utility than does the number of accounts being billed.

The choice of independent variables in each model is first determined by which variables are available. In the commercial and industrial sectors, not only are there few economic variables in publically-available databases, but the types of variables differ. For example, in the same U.S. Bureau of Economic Analysis local area personal income database, county-level number of employees is available annually for disaggregated commercial sector NAICS but not for disaggregated manufacturing sector NAICS. However, county-level industry employee earnings is available at a disaggregate level for both sectors.

A second criteria for choosing independent variables in each model is more complex and contextual. It has to do with the appropriateness of the variable, its explanatory power, and its overlap with similar variables. For example, including a variable representing the number of room in a housing unit in a model of single family home electricity consumption is appropriate and may have reasonably good explanatory power. However, it may also be correlated with household income. On top of that, household income may be correlated with years of schooling. The choice of including one, two, or all three variables in the model is judgmental and ultimately depends on the performance of each variable alone and together, and most importantly, on the goal of the analysis. If the income effect is the primary phenomenon of interest, perhaps the other two variables should be excluded from the model. On the other hand, if the focus is on the energy consumption attributable to extra rooms and extra years of schooling, perhaps it is the income variable that should be excluded.

For each sector, the goal in modeling and analysis is to maintain the same functional form and the same variables. In keeping with model standardization rather than customization, the semi-log functional form is used for all models. This is a conventional form. In practice it means that the dependent variable, *energy use per site*, is transformed into natural logarithms while one or more of the independent variables is left in its original, linear form. The semi-log form works especially well for estimating energy efficiency policy impacts. Later on in this section the method for calculating policy impacts with an linear independent variable will be described.

2.2 Indicators of Policy Impacts

The datasets provided for this project by the IOUs contain census tract-level information on customer energy use. In addition, they contain census tract-level information on energy efficiency products purchased through each IOU's downstream energy efficiency programs. For all three economic sectors, this information includes ex ante energy reduction estimates and the total costs of the energy efficient products. For the commercial and industrial sectors, the data also include the incentives paid by the IOUs to consumers purchasing the energy efficient products. All three of these variables come in to play in one model or another in estimating policy impacts. Ex ante energy consumption reductions and incentive costs are used as policy impact indicators, and total measure costs are used as instruments for these two variables, a subject to be discussed further on.

IOU energy efficiency program estimates of ex ante energy reductions and incentive costs are used as policy impact indicators because their levels tend to be highly correlated with the full scope of energy efficiency policy, and hence policy-related energy reductions. In this study, these values are cumulated year-over-year to produce monotonically increasing values that represent the current year, and all prior year, policy impacts. Letting s be the ex ante annual energy reductions from every individual measure purchased through a downstream program, total energy reduction per cross section and time period, S, is:

(1) $$S_{it} = \text{Total Annual Ex Ante Energy Reduction}_{it} = \sum s_{it}$$

where i represents a cross section and t represents a time period. Letting *IMC* represent incentive costs and *TMC* represent total measure costs, cumulative ex ante energy consumption reductions (*SAVCUM)*, cumulative incentive costs (*INCCUM)*, and cumulative total measure costs, (*MEACUM)*, are calculated as:

(2) $$SAVCUM_{it} = S_{i,t} + \mathrm{S}_{i,t-1} + ... + \mathrm{S}_{i,t-n}$$

(3) $$INCCUM_{it} = IMC_{i,t} + IMC_{i,t-1} + ... + IMC_{i,t-n}$$

(4) $$MEACUM_{it} = TMC_{i,t} + TMC_{i,t-1} + ... + TMC_{i,t-n}$$

The two monetary values, as well as all the monetary values analyzed in this study, are transformed into constant 2010 dollars using the most recent GDP implicit price deflator.

When the range in energy use from cross section to cross section or time period to time period is not miniscule, using the absolute values just defined are likely to be ineffective as explanatory variables. To produce a scaled, relative value, in equation (5) cumulated ex ante reductions in a given cross section and year are divided by total energy use per cross section per year, e_{it}, to produce *SAVCUMRATIO*, or Z_1. This is the amount of ex ante cumulative energy reduction in any year relative to the actual energy consumption in that year, referred to as e_{it}. For scaling purposes, in equation (6) cumulative incentive costs are divided by the total energy expenditures or bills, $bill_{it}$, in a given year, to produce *INCCUMRATIO*, or Z_2, and in equation (7) cumulative total measure costs are divided by $bill_{it}$ to produce *MEACUMRATIO*, or W_1.

(5) $$SAVCUMRATIO_{it} = Z_{1,it} = \sum (S_{i,t} + S_{i,t-1} + ... + S_{i,t-n}) / e_{it}$$

(6) $$INCCUMRATIO_{it} = Z_{2,it} = \sum (IMC_{i,t} + IMC_{i,t-1} + ... + IMC_{i,t-n}) \; / bill_{it}$$

(7) $$MEACUMRATIO_{it} = W_{1,it} = \sum (TMC_{i,t} + TMC_{i,t-1} + ... + TMC_{i,t-n}) / bill_{it}$$

For commercial and industrial sector natural gas consumption, the denominator *bill* in equations (6) and (7) is replaced with total therm consumption, $e_{therm,\,it}$. This is due to the fact that total natural gas expenditures are not available for these sectors for this study.

Even with these cumulative indicators it is often possible that policy impacts remain undetectable. This can be due to the fact that the indicator values, despite being cumulative, are small. For example, a cross section whose expected reductions in each year is 0.5 percent of total energy use might only have a cumulative reduction ratio, Z_1, of 2 percent after four year (this could vary somewhat depending on whether other factors cause energy use in year four to increase or decrease). In such cases, the best hope for detection is to trim the analysis sample so that it only includes those cross sections where detection is possible. There is no rule of thumb to what the cutoff of relative indicator values must be. However, if there is a reasonable point at which policy impacts can be detected, then at the very least it is possible to reject the hypothesis that the program had no effect.

There are several ways to create sample cutoff points. One way is to use increasingly stringent ratios to screen observations, another is to rank observations by ratio and then apply cutoffs by rank, and a third is to select different fractions of the samples based on going from low-to-high or high-to-low ratios. Since all three of these methods are based on the order of the ratio values, they all lead to similar results.

2.3 Instrumental Variables

Despite large differences in the amount, and types, of variables available for modeling the consumption of two fuels in three economic sectors, it is possible to provide a description of the general model specification that will be used for most of the analyses in this study. With p_{it} symbolizing the number of sites or premises in each cross section in each year, annual energy consumption per site per cross section and year, E_{it}, is calculated as

(8) $$E_{it} = \text{Annual Energy Consumption Per Site}_{it} = e_{it} / p_{it}$$

For the multivariate analyses in this study the relationship between E_{it} and a policy impact indicator variable, either Z_1 or Z_2, is characterized by the following two simultaneous equations:

(9) $$E_{it} = a'_0 + a_1 X'_{it} + a_2 Z_{it} + u_{it}$$

(10) $$Z_{it} = b'_0 + b_1 E_{it} + b_2 W'_{it} + v_{it}$$

Equation (9) is what must be estimated in order to measure policy impacts. In it, a'_0 represents one or more constants, a_1 and a_2 represent non-zero coefficients associated with independent variables, and u is the model error term. Also, the vector X' contains independent variables that are causally related to energy use, such as climate, and the variable Z is an indicator of policy impacts. It is the coefficient of this variable, a_2, that expresses the relationship between energy efficiency policy and energy use.

Equation (10) shows that while Z influences E, the reverse is also true. This poses a problem in estimating a_2 in equation (9) because independence of the right-hand variables from the left-hand variable is a necessary condition for regression models to produce consistent, unbiased coefficients. Estimating equation (9) without correcting for the relationship between Z

and E will lead to Z being correlated with u, resulting in an inconsistent and biased estimate of the value of a_2.

Since the primary goal of this study is to investigate the degree to which energy efficiency policy has had an impact on energy use, the endogeneity of Z must be remedied. Fortunately, equation (10) not only points out the problem, but points to the solution. The independent variables in equation (10), denoted by W', can be used to remove the correlation between Z and u. These variables are correlated with Z, but independent of E and are referred to in the context of simultaneous equation estimation as *instruments*.

The technique for solving the simultaneity problem involves first estimating equation (10) and then using the coefficients of this model to produce forecasts of Z, shown as Z^*, in equation (11).

(11) $$Z^*_{it} = b'_0 + b_1 X'_{it} + b_2 W'_{it}$$

These forecasts replace the original values of Z in equation (9), and the new model, shown as equation (12), is then estimated.

(12) $$E_{it} = c'_0 + c_1 X'_{it} + c_2 Z^*_{it} + u^*_{it}$$

This simultaneous equations technique is known as two-stage least squares, TSLS, or as instrumental variables estimation. If all the necessary conditions are met, then unlike a_2, the coefficient c_2 is a consistent estimate of the impact of policy activities on energy use, and a less biased one.

Although theoretically sound, from a practical perspective, there remains the possibility that the TSLS coefficient estimates in equation (12) can be inferior to those produced by ordinary least squares (OLS) estimation in equation (9). This can occur when the Z is, in fact, not endogenous, or when the instruments in the first stage regression, equation (10), are poorly correlated with Z. Diagnostic tests are available to assess these issues.

The energy efficiency policy impact indicator used for the residential and manufacturing models is Z_1, the cumulative ex ante energy reduction ratio. For the commercial sector models the cumulative incentive costs ratio, Z_2 , is used. For both indicators, the same two instruments are employed. The first is previously defined in equation (7) and symbolized by W_1, i.e., the cumulative total measure cost ratio. This variable is not related to energy use, but is likely to be closely correlated with the ex ante energy reductions and incentive costs ratios. The second instrument is energy supply costs, *SC*. This is the unit cost of energy (either electricity or natural gas) per cross section and year. It is calculated as:

$$SC_{it} = W_{2,it} = \frac{bill_{it}}{e_{it}} = \frac{\sum\left(es_{j,it} \times r_{j,it}\right)}{e_{it}} \tag{13}$$

where $es_{j,it}$ is the energy use for each site *j* within a cross sectional unit *i* in year *t*, and $r_{j,it}$ is the IOU rate schedule faced by each site. The sum of this product is the total expenditures on a fuel in a particular cross section and year. Dividing this value by the total energy use in a cross section, e_{it}, results in the unit supply cost, SC_{it},, or $W_{2,it}$. W_2 is also expected to be related to the cumulative ex ante energy reductions ratio (Z_1), and the cumulative incentive costs ratio (Z_2). By construction this is an average supply cost that is only indirectly related to E_{it}. This is because IOU rates are multi-tiered and administered by the CPUC based on costs of service, time of use, and so on. Thus, expenditures can differ substantially between two or more cross sections even when the same amount of total energy is purchased. The rates or prices facing consumers may be identical, but the application of the rates will differ based on patterns of energy use and/or the number of consumers in a cross section. The average cost per unit of energy, *SC*, thus reflects the costs of supply for a pattern of energy use rather than the price for a fixed quantity of energy use.

2.4 Policy Impact Estimation

For this study, the energy reductions attributable to energy efficiency policy are calculated via the coefficient of the policy impact indicator, that is, the coefficient c_2 in equation (12) that is attached either to Z^*_1 or Z^*_2, depending on the specific model. Total cumulative energy reductions (*TCR*) over any model estimation period can then be found by:

(14) $$TCR_{Z_1} = c_{2,Z_1} \times \frac{\sum SAVCUM_t}{\sum e_t} \times \bar{e}_t$$

(15) $$TCR_{Z_2} = c_{2,Z_2} \times \frac{\sum INCCUM_t}{\sum bill_t} \times \bar{e}_t$$

where *SAVCUM* , *INNCUM*, *e*, and *bill* are as previously defined and $\bar{e}$ represents average annual energy consumption per year. As the individual models that produce c_2 only contain samples of the relevant populations (due to missing values, sample trimming, etc.), to calculate the total policy impacts, *TCR*, data for the entire model populations are used to produce *ΣSAVCUM*, *ΣINCCUM*, *Σe, Σbill,* and $\bar{e}$. In plain language, the calculation of policy impacts (*TCR*) using *SAVCUMRATIO* as the *Z* variable can be seen as the coefficient of *Z* (the marginal impact of the policy indicator over the estimation period) multiplied by (a) the ratio of aggregate ex ante reductions in energy use over the estimation period to aggregate energy consumption over the estimation period, multiplied by (b) average annual energy consumption over the estimation period.

It is important to emphasize that these policy impact estimates are calculated for the average annual energy consumption for all the years in the model estimation period, not for any individual year. This is because the coefficient c_2 represents the marginal impact of *Z* on energy consumption over all the years in the estimation period, not any one particular year; in other words, it is an *average* marginal impact for the period as a whole. It follows, for example in equation (14), that the first multiplicand for c_2 is total *SAVCUM* over the entire estimation period divided by total energy consumption over the entire estimation period (not single year total *SAVCUM* divided by the single year total energy consumption). And likewise, the second multiplicand for c_2 is average total energy consumption over the all the years in the estimation period, not any single year's total energy consumption. Calculated this way, *TCR* is interpretable as the full impact of energy efficiency policy on energy consumption over the estimation period. Further, the baseline for measuring the *percentage change in energy consumption due to energy efficiency policy* is simply the average annual energy consumption per year over the estimation period, $\bar{e}$.

Two final issues related to policy impact estimation are the choice of the policy impact estimation period for evaluating the impacts of 2006-2008 programs, and the method used to produce confidence intervals for the combined findings. The first issue arises because of three considerations:

- not all energy efficiency policy actions in a given year are implemented on January 1 of that year -- rather, they are distributed throughout the 12 months;
- energy efficiency policies continue beyond the specified program evaluation period; and,
- to produce long-term estimates of policy impacts it is desirable to have at least 2 years of post-program data.

Choosing the three years of 2006-2008 as the model estimation period has several disadvantages. For one, it reduces the number of observations populating the models by at least one-fourth, if not two-fifths. For another, assuming that half the ex ante energy reductions reported in every year are actually realized in the same year they are reported (ex ante energy reductions are annualized values), the *TCR* for the 2006-2008 period will include the partial reductions for calendar year 2005 policies and only two-and-a-half of the three policy years of interest. And last, the three year estimation period, weakens the interpretation of the findings as long-term policy impacts.

For these reasons, a better estimation period is 2006 though 2009. Using 2009 data in the analysis adds 33 percent more data, permits *TCR* to represent all of the policy impacts in the 2006-2008 policy period (while also including partial impacts from 2005 and 2009), and allows for a more reliable estimate of long term program impacts.

A third alternative is to use the five years from 2006 through 2010 as the estimation period. This has even greater advantages than using the four year period except for the fact that *TCR* then encompasses part of 2005, all of 2009 and part of 2010 policy impacts. Nevertheless, from a long-term policy perspective this is undoubtedly the best model estimation period. For all the analyses that follow, results are reported for the four and five year model estimation periods.

The final element of the policy modeling framework is the method used for combining the separate model findings into statewide totals and calculating confidence intervals. The actual combining of the estimated policy impacts is done by summing, by fuel, for each estimation period. The standard error associated with each sum is calculated by taking the square root of the sum of the squared standard errors of each estimated policy impact. The 90 percent confidence interval for the combined impacts are calculated by multiplying the aggregated standard error by the z-value of 1.645.

3. Impact Evaluation Findings

3.1 Residential Sector

For all the sector and fuel-specific analyses in this study key energy consumption and program tracking data elements were inspected prior to modeling for missing values, seemingly erroneous data or outliers, and high and low end values that might skew the sample statistics or suggest multi-modal distributions. Because census tracts rather than counties are the units of observation in the residential sector, F-tests were conducted to determine whether the data from the different utilities could be pooled.

Table 1 shows the total sample sizes (number of census tracts) for the two utilities for which residential electricity consumption data are available and the two utilities for which natural gas consumption are available.

Table 1: Residential Sector Sample Sizes, by Utility and Fuel, 2006

2006	**Mean**	**Median**	**Max**	**Min.**	**Std. Dev.**	**n**
kWh/Site/Tract						
PG&E	6,948	6,638	14,994	1,215	2,293	2,830
SDG&E	6,247	5,893	14,734	1,617	2,212	677
Therms/Site/Tract						
PG&E	520	468	4,960	105	269	2,930
SDG&E	405	373	2,434	212	152	609

Based on the findings from the complete dataset, electricity consumption per site per census tract was restricted to be between 1,000 and 15,000 kWh and natural gas consumption per site per census tract had to be greater than zero and less than 5,000 therms per year. These restrictions produced minor losses of observations, e.g., in 2006 a total of 34 census tract in PG&E's service territory were dropped from the electricity consumption analysis and 7 from the natural gas consumption analysis. Additional screening was imposed based on the values of energy supply costs and the cumulative ex ante annual reduction ratios. For the former, estimated natural gas supply costs were restricted to be between 10 cents and one dollar (in constant 2010 dollars) and for the latter, the 2009 cumulative ex ante reduction ratio for each fuel could not exceed 50 percent.

Initial diagnostic tests, such as pairwise F-tests of the equality of the electricity consumption model coefficients, indicated that the null hypothesis of no differences between utilities could be rejected. As a result, all of the analyses in the residential sector are performed separately for each utility and each fuel. The PG&E and SDG&E electricity consumption models for the two different estimation periods, 2006 to 2009 and 2006 to 2010, are displayed in Tables 2 and 3, respectively.

Table 2: Residential Sector Electricity Consumption Model, PG&E

PG&E	**Description**	**2006-2009**		**2006-2010**	
lnKWH	kWh per Site/Census Tract	Coeff.	SE	Coeff.	SE
C	Intercept	**7.49318**	0.03829	**7.52415**	0.03895
D07	2007	**0.04328**	0.00923	**0.04907**	0.00989
D08	2008	0.01523	0.01527	0.02718	0.01645
D09	2009	**0.05092**	0.01696	**0.06383**	0.01828
D10	2010			**0.09686**	0.02018
XAGGHHINCOME	Aggregate Income	**0.00000**	0.00000	**0.00000**	0.00000
AVGSIZEHH	Average Household Size	**-0.02958**	0.00505	**-0.02773**	0.00485
COLLEGE	People went to College	**-0.00010**	0.00001	**-0.00009**	0.00001
GROUPPOP	Population in Group Housing	0.00001	0.00001	0.00001	0.00001
MEDIANAGE	Median Age	**0.00162**	0.00049	**0.00194**	0.00049
MEDIANROOMS	Median rooms of HU	**0.13712**	0.00367	**0.13144**	0.00343
HU1DETACHED	Detached Housing Unites	0.00001	0.00001	**0.00001**	0.00001
HU3OR4	3 and 4-plexs	**-0.00032**	0.00002	**-0.00030**	0.00002
HUMOBILEHOME	Mobile Homes	**0.00038**	0.00002	**0.00038**	0.00002
HUBOATRVVAN	Boat, RV, Van Housing Units	0.00000	0.00029	-0.00003	0.00026
BUILT2004DUM	Dum 30% New 2000 to 2004	**-0.01897**	0.00978	**-0.02003**	0.01039
VACANTHU	Number of Vacant HU	**-0.00015**	0.00002	**-0.00017**	0.00002
HDD	Heating Degree Days	**0.00015**	0.00001	**0.00014**	0.00001
CDD	Cooling Degree Days	**0.00025**	0.00000	**0.00024**	0.00000
Z*1	Cum. ex ante kWh Savings Ratio	**-4.55708**	1.09533	**-5.45169**	1.18636
Adj. R-sqd		0.62		0.57	
n		10,700		13,336	
TCR	Total Cum. Reduction (GWh)	1,408		1,923	
% TCR	TCR/Avg. Ann. Consum.	4.7%		6.4%	

Table 3: Residential Sector Electricity Consumption Model, SDG&E

SDG&E	**Description**	**2006-2009**		**2006-2010**	
lnKWH	kWh per Site/Census Tract	Coeff.	SE	Coeff.	SE
C	Intercept	**7.32904**	0.04004	**7.33981**	0.03425
D07	2007	0.01019	0.01179	0.00881	0.01062
D08	2008	0.00256	0.01211	0.00114	0.01075
D09	2009	**0.04659**	0.01819	**0.03967**	0.01489
D10	2010			**0.05508**	0.01859
AGGHHINCOME	Aggregate Income	**0.00000**	0.00000	**0.00000**	0.00000
AVGSIZEHH	Average Household Size	**-0.04592**	0.00824	**-0.04200**	0.00698
COLLEGE	People went to College	**-0.00010**	0.00001	**-0.00010**	0.00001
GROUPPOP	Population in Group Housing	**0.00001**	0.00000	**0.00001**	0.00000
MEDIANAGE	Median Age	0.00070	0.00078	0.00094	0.00066
MEDIANROOMS	Median rooms of HU	**0.23176**	0.00633	**0.22784**	0.00544
HU1DETACHED	Detached Housing Unites	**-0.00006**	0.00001	**-0.00006**	0.00001
HU3OR4	3 and 4-plexs	**-0.00023**	0.00003	**-0.00022**	0.00003
HUMOBILEHOME	Mobile Homes	**0.00038**	0.00003	**0.00038**	0.00002
HUBOATRVVAN	Boat, RV, Van Housing Units	**0.00198**	0.00056	**0.00207**	0.00051
BUILT2004DUM	Dum 30% New 2000 to 2004	**-0.03192**	0.01821	-0.02890	0.01628
VACANTHU	Number of Vacant HU	-0.00004	0.00003	**-0.00005**	0.00002
HDD	Heating Degree Days	**0.00012**	0.00001	**0.00011**	0.00001
CDD	Cooling Degree Days	**0.00013**	0.00001	**0.00013**	0.00001
Z*1	Cum. ex ante kWh Savings Ratio	**-4.60959**	1.24041	**-4.09427**	0.94016
Adj. R-sqd		0.78		0.78	
n		2,697		3,376	
TCR	Total Cum. Reduction (GWh)	290		314	
% TCR	TCR/Avg. Ann. Consum.	3.9%		4.2%	

Variable mnemonics beginning with "ln" refer to their values being transformed into natural logarithms. In these models, as in all the models in this study, the base year for the fixed time effects is 2006.

Hundreds of population and housing-related variables are available via the Census. Nevertheless, the twelve Census variables specified in all the residential models are identical, selected intuitively based on their general appropriateness for explaining residential sector electricity and natural gas consumption. This one-size-fits-all approach leads to models whose coefficients are not always statistically significant, and which do not necessarily produce the very best model for a particular utility or fuel. However, standardized models are best for addressing the CPUC's pilot study goals. Future studies can delve further into optimal model specifications.

For the 2006-2010 estimation period, the PG&E model indicates that the average annual policy impact was 6.4 percent reduction in average annual GWh consumption. This is shown by the *%TCR* statistic, which is formed by dividing estimated total cumulative reduction (*TCR*) by average annual total electricity consumption. In Table 3, the SDG&E model indicates for the 2006-2010 estimation period that the cumulative policy impact was 4.2 reduction in average annual GWh consumption.

The PG&E and SDG&E natural gas consumption models for the two different estimation periods, 2006 to 2009 and 2006 to 2010, are displayed in Tables 4 and 5. For these models, the program impact indicator, $Z_{1,}$ is cumulative ex ante natural gas savings, the instrument W_1 is cumulative natural gas measure costs, and the instrument W_2 is natural gas supply costs, constructed in a similar manner to how residential electricity supply costs were constructed. Endogeneity tests indicate that the null hypothesis of exogeneity can be rejected at the 95 percent confidence level or greater; weak instrument tests indicate that the null hypothesis can also rejected for the hypothesis that the coefficients of the two instruments are jointly zero.

Table 4: Residential Sector Natural Gas Consumption Model, PG&E

PG&E	**Description**	**2006-2009**		**2006-2010**	
lnTHERM	Therms per Site/Census Tract	Coeff.	SE	Coeff.	SE
C	Intercept	**5.82048**	0.05236	**5.85056**	0.04418
D07	2007	-0.01127	0.00764	-0.00582	0.01203
D08	2008	**-0.03518**	0.01112	**-0.01227**	0.00731
D09	2009	-0.00485	0.01373	-0.03418	0.00894
D10	2010			-0.00445	0.01011
XAGGHHINCOME	Aggregate Income	**0.00000**	0.00000	**0.00000**	0.00000
AVGSIZEHH	Average Household Size	**-0.01803**	0.00729	**-0.02345**	0.00617
COLLEGE	People went to College	**-0.00009**	0.00001	**-0.00009**	0.00001
GROUPPOP	Population in Group Housing	-0.00001	0.00001	0.00000	0.00001
MEDIANAGE	Median Age	**0.00763**	0.00086	**0.00727**	0.00072
MEDIANROOMS	Median rooms of HU	0.00857	0.00755	0.01125	0.00656
HU1DETACHED	Detached Housing Unites	**-0.00014**	0.00001	**-0.00014**	0.00001
HU3OR4	3 and 4-plexs	**-0.00040**	0.00004	**-0.00042**	0.00004
HUMOBILEHOME	Mobile Homes	**0.00029**	0.00004	**0.00030**	0.00004
HUBOATRVVAN	Boat, RV, Van Housing Units	0.00044	0.00040	0.00031	0.00034
BUILT2004DUM	Dum 30% New 2000 to 2004	**-0.09152**	0.01244	**-0.09601**	0.01099
VACANTHU	Number of Vacant HU	**0.00013**	0.00004	**0.00015**	0.00003
HDD	Heating Degree Days	**0.00004**	0.00001	**0.00003**	0.00001
CDD	Cooling Degree Days	**-0.00002**	0.00001	**-0.00003**	0.00001
Z*1	Cum. ex ante Therm Savings Ratio	-3.34138	39.52364	-7.34058	25.06951
Adj. R-sqd		0.36		0.38	
n		6,748		8,773	
TCR	Total Cum. Reduction (MDth)	2,398		6,560	
% TCR	TCR/Avg. Ann. Consum.	1.2%		3.3%	

For the 2006-2010 estimation period, the PG&E model indicates that the average annual policy impact was a 3.3 cumulative increase in average annual therm consumption. The SDG&E model indicates for the 2006-2010 estimation period that the cumulative policy impact was a 64.1 percent increase in average annual therm consumption.

Table 5: Residential Sector Natural Consumption Model, SDG&E

SDG&E	**Description**	**2006-2009**		**2006-2010**	
lnTHERM	Therms per Site/Census Tract	Coeff.	SE	Coeff.	SE
C	Intercept	**5.81180**	0.07470	**5.77604**	0.06603
D07	2007	**0.03752**	0.01415	**-0.03065**	0.02264
D08	2008	-0.02361	0.01473	0.03940	0.01399
D09	2009	**-0.08450**	0.01961	**-0.02229**	0.01436
D10	2010			-0.07801	0.01821
AGGHHINCOME	Aggregate Income	**0.00000**	0.00000	**0.00000**	0.00000
AVGSIZEHH	Average Household Size	**-0.06368**	0.01293	**-0.06382**	0.01111
COLLEGE	People went to College	**-0.00021**	0.00002	**-0.00020**	0.00002
GROUPPOP	Population in Group Housing	**0.00003**	0.00001	**0.00003**	0.00001
MEDIANAGE	Median Age	0.00044	0.00106	0.00057	0.00091
MEDIANROOMS	Median rooms of HU	**0.04327**	0.00940	**0.04733**	0.00842
HU1DETACHED	Detached Housing Unites	**-0.00018**	0.00002	**-0.00017**	0.00002
HU3OR4	3 and 4-plexs	**-0.00084**	0.00008	**-0.00081**	0.00007
HUMOBILEHOME	Mobile Homes	**0.00025**	0.00003	**0.00026**	0.00003
HUBOATRVVAN	Boat, RV, Van Housing Units	-0.00026	0.00063	-0.00047	0.00065
BUILT2004DUM	Dum 30% New 2000 to 2004	-0.05132	0.04349	-0.04072	0.03939
VACANTHU	Number of Vacant HU	**0.00031**	0.00007	**0.00035**	0.00007
HDD	Heating Degree Days	**0.00004**	0.00001	**0.00004**	0.00001
CDD	Cooling Degree Days	**0.00008**	0.00002	**0.00009**	0.00001
Z*1	Cum. ex ante Therm Savings Ratio	**176.15380**	47.60054	**155.14510**	42.47758
Adj. R-sqd		0.37		0.37	
n		2,213		2,806	
TCR	Total Cum. Reduction (MDth)	-15,912		-19,859	
% TCR	TCR/Avg. Ann. Consum.	-51.3%		-64.1%	

3.2 Commercial Sector

Table 6 lists the 2-digit NAICS (North American Industry Classification System) codes for the commercial sector and their recoded classifications after combining industries. Table 7 contains summary statistics related to kWh per site per county for those counties in which the kWh per site in an industry is more 20,000 kWh and less than 1,00,000 kWh per year. In 2006, these cutoffs result in a loss on the low end of 83 county observations (8 of which had kWh per site values of zero and 42 of which had values that were greater than zero but less than 10,000 kWh). On the high end, 4 county observations were lost.

Table 6: Commercial Sector NAICS and Study Codes

2-digit NAICS	Industry Description	Study Recode
51	Information	C1
52	Finance and insurance	(Office)
53	Real estate and rental and leasing	
54	Professional and technical services	
55	Management of companies and enterprises	
92	Government and government enterprises	
44,45	Retail Trade	C2
61	Educational services	C3
62	Health care and social assistance	C4
71	Arts, entertainment, and recreation	C5
72	Accommodation and food services	C6
42	Wholesale trade	C7
48,49	Transportation and warehousing	(Misc.)
56	Administrative and waste services	
81	Other services, except public administration	

Table 7: Commercial Sector Annual kWh per Site, 2006

Code	Mean	Median	Max	Min.	Std. Dev.	n
C1	66,369	51,921	278,945	20,046	54,488	56
C2	123,386	125,572	212,129	23,050	42,630	60
C3	219,141	199,341	743,494	23,957	120,560	56
C4	101,401	98,541	188,140	32,990	39,464	52
C5	105,548	79,870	727,026	24,195	107,843	54
C6	110,827	113,251	244,193	23,089	44,463	60
C7	58,018	49,650	185,885	20,811	34,979	45
All	113,902	97,076	743,494	20,046	86,245	383

In the design of this study all the energy consumption and program tracking data for the commercial and industrial sectors are built up from census tracts to counties within each utility service territory. As such, some counties are entered into the analysis multiple times, because some of the counties in California are served by more than one IOU. Table 8 contains the same information for the sample used for the econometric analysis. It consists of roughly one-third of the observations of the larger sample, the ones with the higher cumulative ex ante electricity reduction ratios, Z_1, in 2009. These observations were selected because the preliminary analyses were unable to detect energy reductions from the original sample. Note that the relative sample sizes of the industries change, but the electricity consumption statistics for each industry are roughly similar.

Table 8: Restricted Sample Commercial Sector Annual kWh per Site, 2006

Code	Mean	Median	Max	Min.	Std. Dev.	n
C1	56,645	66,102	72,361	31,473	22,023	3
C2	124,558	126,843	188,622	48,988	33,889	41
C3	250,399	226,985	445,547	125,787	93,264	14
C4	90,306	81,357	152,262	64,918	31,867	6
C5	69,773	51,636	147,513	35,195	41,974	7
C6	106,036	108,545	176,894	38,029	31,378	34
C7	57,901	47,461	185,885	23,803	34,755	30
All	112,258	105,549	445,547	23,803	68,637	135

Preliminary analysis of the commercial sector models revealed that even after restricting the sample based on the Z_1 rankings, the variable itself exhibited a high positive correlation with energy use, one that could not be remedied with TSLS. Thus, the alternative indicator of program impacts, the cumulative incentive costs ratio, or Z_2, was included in the specification. As this is a financial ratio, no restrictions are placed on the magnitude of this value. The findings for the models containing Z_2, are shown in Table 9. As before, variable mnemonics beginning with "ln" refer to their values being transformed into natural logarithms.

In this model, the base for the fixed cross section effect for utilities is SDG&E and the base for fixed industry effects is the collection of miscellaneous industries , C7. It is important to note that by employing fixed cross section effects by industry and utility, it is not possible to include county fixed effects in the model. This is not a cause for concern because counties within a single state in and of themselves are not likely to play a major role in the influencing the energy use of any industries. However, county-level weighting, by industry mix, is used to control for county-level heteroscedasticity. For each model, county-specific residuals vectors are used to form county-specific variances, and then weighted least squares (WLS) is applied to form feasible GLS estimates.

In addition to the policy impact indicator, the continuous variables in the model represent industry earning per employee, the share total county employees that are in each industry, and heating and cooling degree days. The findings indicate that for either estimation period, the coefficient of the policy impact indicator, Z_2, is not statistically significant. Endogeneity testing indicates that the null hypothesis of exogeneity can be rejected for the 2006-2010 period but not for the 2006-2009 period. Weak instrument testing indicates that the hypothesis that the two instrument coefficients, W_1 and W_2, are jointly zero can be rejected. The 2006-2010 estimation period findings indicate that cumulative policy impacts for this sample, which includes all three IOUs, are 2.1 percent of average annual electricity consumption.

Table 9: Commercial Sector Electricity Consumption Model

PG&E, SDG&E, SCE	**Description**	**2006-2009**		**2006-2010**	
lnKWH	kWh per Site/Industry/County	Coeff.	SE	Coeff.	SE
C	Intercept	**14.536**	0.400	**14.704**	0.359
C1	Office	0.065	0.071	0.078	0.071
C2	Retail Trade	**1.040**	0.028	**1.040**	0.023
C3	Educational Services	**1.935**	0.104	**1.935**	0.090
C4	Health Care	**0.708**	0.038	**0.716**	0.033
C5	Arts and Entertain.	**0.833**	0.063	**0.804**	0.054
C6	Accommodation, Food	**1.013**	0.040	**0.993**	0.035
PGE	Fixed CS (Utility) Effect	**0.576**	0.078	**0.536**	0.070
SCE	Fixed CS (Utility) Effect	**0.285**	0.076	**0.237**	0.069
D07	Fixed TS (Year) Effect	-0.009	0.018	-0.005	0.016
D08	Fixed TS (Year) Effect	0.034	0.025	0.036	0.022
D09	Fixed TS (Year) Effect	-0.007	0.033	-0.004	0.026
D10	Fixed TS (Year) Effect			0.008	0.032
lnXEARNPEREMP	Per Emp. Earning/Ind./Cty.	**0.360**	0.047	**0.335**	0.042
lnEMPSHARE	Share of Employs/Indust./Cty.	**0.072**	0.033	**0.075**	0.029
lnHDD	Heating Degree Days	**-0.675**	0.037	**-0.680**	0.033
lnCDD	Cooling Degree Days	**-0.022**	0.004	**-0.019**	0.003
Z^*_2	Cum. Incent. Cost Ratio	-0.418	0.519	-0.552	0.370
Adj. R-sqd		0.92		0.93	
n		540		675	
TCR	Total Cum. Reduction (GWh)	917		1,612	
% TCR	TCR/Avg. Ann. Consum.	1.2%		2.1%	

Table 10 contains summary statistics related to therms per site per county. Inspection of the full dataset led to a lower limit cutoff of 1,000 therms per site per industry per county, and an upper limit cutoff of 10,000. Of the non-zero valued therms per site counties, in 2006 there were 80 sites that were below the lower cutoff and 16 were above the upper cutoff.

Table 10: Commercial Sector Therms per Site, 2006

Code	Mean	Median	Max	Min.	Std. Dev.	n
C1	3,304	3,214	7,054	1,001	2,000	20
C2	1,914	1,385	5,639	1,004	1,293	13
C3	5,715	5,331	9,697	1,054	2,470	24
C4	4,352	3,875	9,575	1,517	2,181	28
C5	3,151	2,473	7,586	1,011	1,790	21
C6	4,384	4,548	8,611	1,587	1,811	33
C7	2,301	1,687	6,457	1,132	1,517	24
All	3,779	3,344	9,697	1,001	2,234	163

Table 11 contains the estimated model and policy findings for commercial sector natural gas consumption, where all the independent variables and all the estimation procedures are the same as for the electricity consumption models. Note that estimates of natural gas supply costs are not available for this study, hence electricity supply cost are used as an instrument in lieu of natural gas supply costs. The 2006-2010 estimation period findings indicate that cumulative policy impacts are 19.5 percent of average annual natural gas consumption. Unlike for the coefficient of the electricity consumption model, for this model the policy impact indicator, Z_2, is highly statistically significant. Endogeneity testing indicates that the null hypothesis of exogeneity can be rejected at the 95 percent level for both periods. Weak instrument testing indicates that the hypothesis that the two instrument coefficients, W_1 and W_2, are jointly zero can be rejected.

Table 11: Commercial Sector Natural Gas Consumption Model

PG&E, SDG&E	**Description**	**2006-2009**		**2006-2010**	
lnTherms	Therms per Site/Industry/County	Coeff.	SE	Coeff.	SE
C	Intercept	**9.849**	0.903	**9.685**	0.791
C1	Office	**-0.586**	0.118	**-0.595**	0.103
C2	Retail Trade	**-0.528**	0.076	**-0.509**	0.068
C3	Educational Services	**1.572**	0.139	**1.645**	0.121
C4	Health Care	**0.304**	0.106	**0.328**	0.095
C5	Arts and Entertain.	**0.731**	0.132	**0.783**	0.117
C6	Accommodation, Food	**0.995**	0.092	**1.045**	0.079
PGE	Fixed CS (Utility) Effect	**-0.585**	0.097	**-0.607**	0.086
SCE	Fixed CS (Utility) Effect				
D07	Fixed TS (Year) Effect	**0.159**	0.044	**0.142**	0.043
D08	Fixed TS (Year) Effect	**0.327**	0.053	**0.296**	0.048
D09	Fixed TS (Year) Effect	**0.370**	0.060	**0.335**	0.053
D10	Fixed TS (Year) Effect			**0.395**	0.063
lnXEARNPEREMP	Per Emp. Earning/Ind./Cty.	**0.475**	0.060	**0.497**	0.052
lnEMPSHARE	Share of Employs/Indust./Cty.	**0.216**	0.056	**0.225**	0.051
lnHDD	Heating Degree Days	**-0.335**	0.100	**-0.328**	0.087
lnCDD	Cooling Degree Days	-0.018	0.012	-0.012	0.010
Z^*_2	Cum. Incent. Cost Ratio	**-40.979**	7.313	**-34.477**	5.550
Adj. R-sqd		0.68		0.70	
n		689		864	
TCR	Total Cum. Reduction (MDth)	222.478		234.512	
% TCR	TCR/Avg. Ann. Consum.	18.6%		19.5%	

3.3 Industrial Sector

Broadly speaking, the industrial sector of the U.S. economy is made up of natural resources, construction, and manufacturing industries. Table 12 lists the 2 or 3-digit NAICS codes associated with these industries, a brief description of the industries the codes represent, and the consolidation and recoded classification of industries developed for this study. The industry data consolidation scheme is based on practical considerations. Industry I4 combines 2 industries that are frequently combined in government statistics, and industry code I12 combines 12 industries that individually do not account for a large proportion of California's industrial electricity use and/or have few observations.

Table 12: Industrial Sector NAICS and Study Codes

NAICS	Industry Description	Study Codes
11	Agriculture, Forestry, Fishing and Hunting	I1
21	Mining, Quarrying, and Oil and Gas Extraction	I2
22	Utilities	I3
312 311	Beverage and Tobacco Product Manufact. Food Manufact.	I4
334	Computer and Electronic Product Manufact.	I5
327	Nonmetallic Mineral Product Manufact.	I6
324	Petroleum and Coal Products Manufact.	I7
326	Plastics and Rubber Products Manufact.	I8
333	Machinery Manufact.	I9
325	Chemical Manufact.	1I0
321	Wood Product Manufact.	I11
332 322 336 331 323 339 335 337 315 316 314 313	Fabricated Metal Product Manufact. Paper Manufact. Transportation Equipment Manufact. Primary Metal Manufact. Printing and Related Support Activities Miscellaneous Manufact. Elec. Equip., Appli. and Component Manufact. Furniture and Related Product Manufact. Apparel Manufact. Leather and Allied Product Manufact. Textile Product Mills Textile Mills	I12
23	Construction	I13

Table13 contains summary statistics related to kWh per site per county. As with the commercial sector, these data are built up from census tracts to counties within each utility service territory. To bring the means and variances in energy use down to reasonable sizes, those counties in which the kWh per site in an industry is less than 10,000 kWh annually, or more than 750,000 kWh annually, are dropped. In 2006, they result in a loss on the low end of 71 county observations (20 of which had kWh per site values of zero), and a loss on the high end of 133 county observations. Raising the high end restriction to 1,000,000 kWh per site hardly affects the attrition rate and only adds 23 additional observations, but produces substantially larger relative variances and distribution skewness. As can be seen in Table 12, even after the cutoffs are applied, the standard deviations in energy use, by industry, are roughly as large as their means.

Table 13: Industrial Sector Annual kWh per Site, 2006

Code	Mean	Median	Max	Min.	Std. Dev.	n
I1	54,607	40,880	185,627	10,086	39,802	58
I2	139,353	58,001	731,521	11,678	196,388	14
I3	112,451	78,719	563,629	10,786	98,581	65
I4	246,418	177,107	714,273	12,125	190,769	30
I5	229,499	165,283	687,077	22,876	194,893	25
I6	140,026	95,079	612,900	14,145	136,248	39
I7	395,612	466,680	747,886	16,966	231,833	17
I8	243,310	213,595	690,533	14,440	198,632	25
I9	211,411	182,200	715,847	13,512	172,055	41
I10	242,270	111,257	747,723	11,900	263,273	30
I11	226,680	144,990	738,069	14,727	211,506	36
I12	197,604	128,880	715,849	14,076	191,068	53
I13	25,540	20,710	84,482	10,367	16,044	47
All	165,826	93,293	747,886	10,086	182,866	480

Table 14: Industrial Sector Therms per Site, 2006

Code	Mean	Median	Max	Min.	Std. Dev.	n
I1	10,499	6,408	28,315	5,503	8,520	7
I2	24,714	14,333	49,404	10,405	21,472	3
I3	127,023	61,970	490,120	10,346	143,296	19
I4	104,744	70,627	457,049	5,969	114,152	25
I5	29,736	18,733	83,426	8,256	24,407	13
I6	66,316	39,939	226,925	5,337	73,130	15
I7	234,504	230,177	465,093	8,839	145,240	14
I8	43,530	41,928	125,410	8,077	32,636	15
I9	26,682	12,248	114,609	6,195	36,333	8
I10	123,765	61,005	431,239	12,500	134,162	15
I11	58,312	24,518	308,565	7,241	85,507	15
I12	43,432	24,669	162,220	6,495	47,029	16
I13	7,510	7,510	7,510	7,510	na	1
All	56,640	28,109	226,925	5,337	60,241	150

Table 14 contains summary statistics related to therms per site per county. All of the features of the industries and the electricity use statistics are identical for natural gas use. In so far as cutoff values are concerned, inspection of the full dataset led to a lower limit cutoff of 5,000 therms per site per industry per county, and an upper limit cutoff of 500,000 therms per site. In 2006, they result in a loss on the low end of 334 county observations (127 of which had therms per site values of zero), and a total loss on the high end of 23 observations.

Table 15 contains the estimated models for industrial sector electricity consumption. The single economic variable in the model is the BEA's estimate of total annual earnings for employees in an industry and county.

Table 15: Industrial Sector Electricity Consumption Model

PG&E, SDG&E, SCE	**Description**	**2006-2009**		**2006-2010**	
lnKWH	kWh per Site/Industry/County	Coeff.	Std. Err.	Coeff.	Std. Err.
C	Intercept	**13.703**	0.805	**13.500**	0.713
I1	Ag. and Forestry	**-2.612**	0.057	**-2.475**	0.085
I2	Oil-Gas Extraction	**-0.431**	0.222	-0.734	0.291
I3	Utilities	**-1.867**	0.057	**-1.747**	0.086
I4	Food, Bev., Tobacco	**-0.821**	0.082	**-0.723**	0.096
I5	Computers and Electronics	**-1.284**	0.057	**-1.141**	0.086
I6	Nonmetallic Minerals	**-1.538**	0.048	**-1.398**	0.079
I8	Plastics and Rubber	**-0.487**	0.078	**-0.381**	0.108
I9	Machinery	**-1.266**	0.054	**-1.129**	0.083
I10	Chemicals	**-1.310**	0.073	**-1.121**	0.096
I11	Wood	**-1.220**	0.070	**-1.103**	0.095
I12	All Other	**-1.752**	0.066	**-1.640**	0.092
I13	Construction	**-4.203**	0.071	**-4.095**	0.093
PGE	Fixed CS (Utility) Effect	**0.884**	0.106	**0.824**	0.079
SCE	Fixed CS (Utility) Effect	**0.900**	0.082	**0.831**	0.068
D07	Fixed TS (Year) Effect	0.047	0.026	**0.061**	0.027
D08	Fixed TS (Year) Effect	**0.124**	0.043	**0.140**	0.036
D09	Fixed TS (Year) Effect	**0.155**	0.071	**0.184**	0.054
D10	Fixed TS (Year) Effect			**0.251**	0.072
lnXEARN	Total Earning/Industry/County	**0.242**	0.014	**0.246**	0.013
lnHDD	Heating Degree Days	**-0.451**	0.094	**-0.440**	0.084
lnCDD	Cooling Degree Days	**-0.042**	0.018	**-0.042**	0.014
Z^*_1	Cum. kWh Savings Ratio	**-6.452**	2.355	**-6.561**	1.576
Adj. R-sqd		0.62		0.60	
n		1,509		1,886	
TCR	Total Cum. Reduction (GWh)	5,668		7,471	
% TCR	TCR/Avg. Ann. Consum.	13.6%		17.8%	

To control for unobserved geographic effects of one kind or another that may affect energy use, heating and cooling degree day are kept as independent variables in the industrial sector models. These are not expected to affect energy use in the usual manner, such as for space conditioning in residential and commercial buildings, so interpretation of their coefficients is problematic. To ensure data quality, a small number of observations were screened out of the model if the value of Z_1, the ratio of cumulative ex ante electricity reductions to annual energy consumption was greater than 75 percent, or if W_2, electricity supply cost, was greater than one dollar. The base for the fixed cross section effect for utilities is SDG&E and the base for the fixed industry effects is I7, petroleum and coal product manufacturing. County-level weighting, by industry mix, is used to control for county-level heteroscedasticity. Endogeneity testing indicates that the null hypothesis of exogeneity can be rejected at greater than the 95 percent confidence level for all four models. Weak instrument testing indicates that the hypothesis that the two instrument coefficients, W_1 and W_2, are jointly zero can be rejected. For the 2006-2010 period cumulative policy impacts were 17.8 percent of annual average electricity use.

Table 16 contains the estimated model and policy findings for industrial sector natural gas consumption, for which data have been made available for this study only for PG&E and SDG&E. As with the prior model, to ensure data quality, observations were screened out Z_1 was greater than 75 percent or W_2, was greater than one dollar. And as with the commercial natural gas consumption model, electricity supply costs were used in lieu of natural gas supply costs.

For the 2006-2010 estimation period the findings indicate that policy impacts were 1.4 percent of annual industrial sector natural gas consumption. Endogeneity testing indicates that the null hypothesis of exogeneity can be rejected at close to the 95 percent level for the 2006-2010 period, but not for the 2006-2009 period. Weak instrument testing indicates that the hypothesis that the two instrument coefficients, W_1 and W_2, are jointly zero can be rejected.

Table 16: Industrial Sector Natural Gas Consumption Model

PG&E, SDG&E	**Description**	**2006-2009**		**2006-2010**	
lnTherms	Therms per Site/Industry/County	Coeff.	Std. Err.	Coeff.	Std. Err.
C	Intercept	-1.105	2.493	1.131	2.238
I1	Ag. and Forestry	**-3.456**	0.084	**-3.449**	0.073
I2	Oil-Gas Extraction	**-3.164**	0.257	**-3.117**	0.224
I3	Utilities	**-2.264**	0.108	**-2.334**	0.088
I4	Food, Bev., Tobacco	**-1.976**	0.105	**-1.989**	0.085
I5	Computers and Electronics	**-3.305**	0.145	**-3.312**	0.120
I6	Nonmetallic Minerals	**-1.927**	0.090	**-1.950**	0.074
I8	Plastics and Rubber	**-2.213**	0.112	**-2.227**	0.106
I9	Machinery	**-3.389**	0.111	**-3.426**	0.091
I10	Chemicals	**-1.794**	0.128	**-1.817**	0.102
I11	Wood	**-2.681**	0.108	**-2.598**	0.091
I12	All Other	**-2.905**	0.126	**-2.962**	0.113
I13	Construction	**-4.469**	0.169	**-4.059**	0.489
PGE	Fixed CS (Utility) Effect	**-1.185**	0.280	**-0.953**	0.258
SCE	Fixed CS (Utility) Effect				
D07	Fixed TS (Year) Effect	0.080	0.045	0.071	0.050
D08	Fixed TS (Year) Effect	0.029	0.037	0.043	0.038
D09	Fixed TS (Year) Effect	0.078	0.047	**0.093**	0.043
D10	Fixed TS (Year) Effect			0.093	0.054
lnXEARN	Total Earning/Industry/County	**0.191**	0.032	**0.195**	0.028
lnHDD	Heating Degree Days	**1.467**	0.275	**1.171**	0.255
lnCDD	Cooling Degree Days	**0.207**	0.075	**0.183**	0.065
Z^*_1	Cum. Therm Savings Ratio	**-1.405**	0.276	**-1.586**	0.233
Adj. R-sqd		0.47		0.42	
n		498		622	
TCR	Total Cum. Reduction (MDth)	56.172		72.730	
% TCR	TCR/Avg. Ann. Consum.	1.1%		1.4%	

3.4 State-Level Findings

The individual policy impact findings from the eight electricity consumption models and the eight natural gas consumption models can be added together to produce state-level estimates of policy impacts.

Over and above the total impacts of energy efficiency policy – which includes the impacts of all old and existing building codes and appliance standards – additional impacts can be calculated for just those housing units built between 2000 and 2004. This can be done by multiplying the coefficient of the variable “BUILT20DUM” in the residential sector models by the annual average energy consumption of the housing units built in the IOU census tracts between 2000 and 2004. To approximate this level of consumption, the number of housing units in each census tract built in this time period was multiplied by the average annual energy use per site in each census tract. The results of this calculation, as well as of the combining of all the model findings, are contained in Tables 16 and 17. As the energy consumption for the housing units built between 2000 and 2004 are already contained in the residential total annual consumption figures, these are not added to the statewide grand totals. Further, ex ante estimates of expected reductions are not available.

The upper panels of Table 17 and 18 contain the policy impact findings for the four year period from 2006-2009; the lower panels contains the findings for the five year period from 2006-2010.

Table 17: State-Level Policy Impact Findings, Electricity

kWh: 2006-2009 Period	**Avg. Ann. Energy Consum.**	**Last Yr. Cum. Ex Ante Savings**	**Pct. Est. Impact**	**Annual Est. Impact**
Ind. (PG&E, SDG&E, SCE)	41,554,138,628	2,048,456,628	13.6%	5,668,443,641
Com. (PG&E, SGG&E, SCE)	76,735,020,087	3,403,092,575	1.2%	916,539,984
Res. (PG&E)	30,132,043,300	480,430,254	4.7%	1,408,088,335
Res. (SDG&E)	7,483,267,512	105,504,485	3.9%	289,567,961
Res. Codes (PG&E)[1]	2,571,722,287		1.9%	48,775,285
Res. Code (SDG&E)[1]	742,521,437		3.2%	23,700,542
Total	155,904,469,527	6,037,483,943		8,355,115,748
Percent Impact		3.9%		5.4%
Pct. Standard Error (+/-)				28.6%
kWh: 2006-2010 Period				
Ind. (PG&E, SDG&E, SCE)	41,879,937,508	2,512,486,691	17.8%	7,471,169,145
Com. (PG&E, SGG&E, SCE)	76,829,480,556	4,412,722,613	2.1%	1,611,926,492
Res. (PG&E)	30,207,548,725	526,324,700	6.4%	1,923,104,970
Res. (SDG&E)	7,432,945,519	129,427,756	4.2%	313,568,580
Res. Codes (PG&E)[1]	2,552,293,975		2.0%	51,114,791
Res. Code (SDG&E)[1]	729,653,042		2.9%	21,086,973
Avg. Total. Ann. Consum.	156,349,912,308	7,580,961,761		11,391,970,952
Percent Impact		4.8%		7.3%
Pct. Standard Error (+/-)				18.9%

[1] Average annual energy consumption for census tracts with >30% houses built 2000-2004 .

Based on the collected findings of all the eight electricity consumption econometric models estimated for this study, in 2009 the total impact of electricity efficiency policy in all sectors, and including residential building code impacts for housing units built between 2000 and 2004, is a decline in energy use 8,355 GWh. This is a 5.4 percent decline relative to the average, total energy consumption per year in the 2006-2009 period. The relative standard error of the impact estimate is 28.6 percent; at the 90 percent confidence level the relative standard error is plus or minus 47 percent. By comparison, in 2009 the cumulative ex ante estimate of electricity reductions for the 2006-2009 period due to downstream, IOU-implemented energy efficiency programs is 3.9 percent relative to 2009 total electricity consumption. It is not the purpose of this study to speculate as to why there are differences in the ex ante and model-based energy consumption reduction estimates. As previously discussed, the cumulative ex ante estimates of energy reductions due to downstream, IOU-implemented energy efficiency programs are used in this study as *indicators* of the impacts of the broader set of public initiatives that comprise *de facto* state-wide energy efficiency *policy*.

Cumulative policy impacts for the 2006-2010 period are 7.3 percent. The relative standard error of this estimate is plus or minus 18.9 percent, or 31 percent at the 90 percent confidence level. The cumulative IOU energy efficiency program ex ante energy reduction estimate is 4.8 percent of average total energy consumption over the five year period.

Table 18: State-Level Policy Impact Findings, Natural Gas

Therms: 2006-2009 Period	**Avg. Ann. Energy Consum.**	**Last Yr. Cum. Ex Ante Savings**	**Pct. Est. Impact**	**Annual Est. Impact**
Ind. (PG&E,SDG&E)	5,124,383,847	87,353,504	1.1%	56,171,929
Com. (PG&E, SDG&E)	1,198,857,181	17,121,520	18.6%	222,477,711
Res. (PG&E)	2,012,166,142	10,406,303	1.2%	23,976,294
Res. (SDG&E)	297,293,772	2,039,215	-51.3%	-152,541,395
Res. Codes (PG&E)[1]	152,939,950		9.2%	13,873,030
Res. Code (SDG&E)[1]	25,880,855		5.1%	1,328,257
Total	8,632,700,942	116,920,542		165,285,826
Percent Impact		1.4%		1.9%
Pct. Standard Error (+/-)				175%
Therms: 2006-2010 Period				
Ind. (PG&E,SDG&E)	5,143,530,663	111,485,401	1.4%	72,730,412
Com. (PG&E, SDG&E)	1,200,231,263	23,026,253	19.5%	234,511,830
Res. (PG&E)	2,018,224,763	13,930,364	3.3%	65,597,381
Res. (SDG&E)	308,983,896	2,778,731	-64.3%	-198,590,214
Res. Codes (PG&E)[1]	152,601,314		9.6%	14,650,642
Res. Code (SDG&E)[1]	25,675,625		4.1%	1,045,537
Total	8,670,970,586	151,220,748		189,945,588
Percent Impact		1.7%		2.2%
Pct. Standard Error (+/-)				244%

[1] Average annual energy consumption for census tracts with >30% houses built 2000-2004 .

Table 18 indicates that the eight natural gas consumption econometric models yield findings of 1.9 percent policy impacts over the four year estimation period. The relative standard error of this estimate is very large, at 175 percent. The IOU energy efficiency program ex ante estimate of natural gas reductions for the four year period is 1.4 percent relative to average total electricity consumption. In the five year period, the impact estimate for natural gas consumption energy efficiency policy rises to 2.2 percent, again with a very large relative standard error.

4. Recommendations

This study demonstrates that a well-founded econometric framework, coupled with an appropriate, large-sample database, can be developed to evaluate the aggregate impact of the 2006-2008 energy efficiency programs on energy consumption. In doing so, it demonstrates that aggregate econometric models employing large samples are capable of accurately measuring the impact of the Commission's energy efficiency efforts on overall electricity and natural gas consumption in California in the context of post-2012 EM&V activities. The potential for accurate measurement is demonstrated by the standard errors that accompany the estimated electricity policy impacts for the 2006-2010 period. No other type of evaluation study can produce a relative error bound of 31 percent (at the 90 percent confidence level) around a state-level policy impact estimate that embraces all three non-transportation sectors of the economy and incorporates the uncertainties due to free ridership, spillover, rebound, measure interaction and retention, behavioral changes, and general economic conditions.

This study also provides material for discussing two further goals articulated by Decision (D.) 10.10.33. For example, it shows that this evaluation approach is likely to be valuable for improving estimates of aggregate reductions in Greenhouse Gases (GHG) emissions from efficiency programs as required in AB32. Also, that it is likely that they can prove valuable for more directly aligning and integrating energy efficiency program findings into the California Energy Commission's (CEC) demand forecasts, and ultimately, the CPUC's resource procurement process. Further assessment of these issues is beyond the scope of this particular effort; the Commission has contract with another party to examine these goals.

General recommendations for integrating this evaluation approach into the permanent portfolio of post-2012 EM&V activities fall into two categories, database development and econometric analysis. They include:

- Expand the database with additional variables, particularly in the commercial and industrial sectors.
- Upgrade the database for easier access and creating a website with the capability to download customized data requests.
- Develop standardized, automated routines for cleaning customer billing and program tracking data;
- Use geographic information system software for collecting and processing local area data.
- Explore the properties of different types of econometric impact estimators.
- Experiment with customized models for different fuels and sectors.
- Analyze various market segments and customer grouping.
- Investigate the possibilities for developing econometric models that target specific programs and policies.

Finally, it is worth noting that with the extensive deployment of smart meters throughout the IOU service territories, it should be possible to derive census tract-level hourly load shapes data for different end uses, equipment, and customer-types. This raises the possibility of using the same modeling approaches and datasets for analyzing hourly demand as are currently being used for annual energy consumption and the impact of energy efficiency policies. If successful, census tract-level analyses of hourly demand can prove to be these invaluable for planning and evaluating load control programs of all kinds.

APPENDIX A: Publically-Available Data Sources and Data Development

A1. Background

Demand Research's Macro Consumption Metrics pilot study explores an alternative way of collecting data and creating market-related independent variables for energy efficiency policy evaluation. It involves gathering together publically-available data produced mainly by the federal government, but also by additional private and public sources. The main advantages of publically-available data are that they are free or inexpensive; they cover geographically-based populations or large, unbiased samples from these populations; and, they span many past years and will continue on into the indefinite future.

This appendix documents the collection of all of the relevant publically-available longitudinal datasets needed for energy efficiency policy evaluation, and the procedures used to combine them. These procedures are technical and focus on normalizing spatial and temporal misalignments so that the variables selected for evaluating the effects of energy efficiency policies in California's IOU service territories are as consistent and error-free as possible. In the next section, the datasets and data collection activities are described. This is followed by a section documenting all of the major data processing and data imputation steps applied to the datasets prior to econometric modeling and statistical estimation.

A2. Data Collection and Initial Processing

Table A1 lists, by economic sector, the sources of publically-available data used for this impact evaluation pilot project. All of these data reside on federal government agency websites.

Table A1: Data Sources for Pilot Project Variables, by Sector

Data Source	Dataset/Variables
All Sectors	
Census Bureau (Census)	Selected datasets and variables
NCDC	Weather station HDD and CDD
Bureau of Labor Statistics (BLS)	Selected datasets and variables
Bureau of Economic Analysis (BEA)	Selected datasets and variables
Federal Reserve Board (FRB)	Selected datasets and variables
Residential Sector, Only	
Census Bureau (Census)	Census of Population and Housing American Community Survey
Commercial and Industrial Sectors, Only	
Bureau of Economic Analysis (BEA)	Local Area Personal Income Accounts

Residential Census and ACS Data

Of the datasets listed, the ones that require the most attention are the Census of Population and Housing (Census) and the American Community Survey (ACS), the latter drawing on survey responses from about 2 million households. Not only are there tens of thousands of variables in these datasets, but it is from boundaries in these datasets that IOU service territory boundaries are defined. In addition, weather stations are assigned to local areas, and the statistics from different layers of geography are put into conformance.

The Census Bureau maintains an FTP site that contains data files for Census 2000 and 2010 census tracts and counties. As can be seen in Table A2 which provides the counts of the areas covered in California for the Census and ACS datasets, in California there are 58 counties whose boundaries that do not change from decade to decade. However, census tracts change for political and demographic reasons, and so in 2010 there were more than a thousand extra census tract in California compared to 2000. This same FTP site also makes available the ACS data files. The ACS data files from 2005 and 2006 contain 1-year estimates of all housing and population variables for all counties and defined areas (such as metropolitan areas) with populations above 65,000 persons; in California, this includes 40 of the 58 California counties. Beginning in 2007, ACS also provides3-year averages (the current year plus the two prior years) for all counties and defined areas with populations greater than 20,000 persons; this includes 51 of California's counties. Lastly, beginning in 2009, 5-year averages are provided for all variables for all census tracts and counties.

Table A2. Geographic Coverage of Census and ACS Data for California

Data Sources	1 Year Tracts	1 Year Counties	3 Year Average Tracts	3 Year Average Counties	5 Year Average Tracts	5 Year Average Counties
Census 2000 (SF1, SF3)	7,049	58	--	--	--	--
ACS 2005	--	40	--	--	--	--
ACS 2006	--	40	--	--	--	--
ACS 2007	--	40	--	51	--	--
ACS 2008	--	40	--	51	--	--
ACS2009	--	40	--	51	7,049	58
ACS 2010	--	40	--	51	8,057	58
Census 2010 (SF1, only)	8,057	58	--	--	--	--

Although many of the procedures used to obtain and process the Census and ACS data files are similar, the data file features vary enough from year to year that each year's data files must be processed individually. The first step in the process is the downloading from the FTP site the file containing all the data for a given year. For example the ACS 2006 file is a 57 MB file that contains 145 separate ASCII comma-delimited text files. A geographic header record file, *G20061ca.txt*, is included in this download, containing geographic links to the data files. Also in the download is a file that contains the data dictionary and the ACS 2006 Production Summary File: Technical Documentation (*ACS_2006_SF_Tech_Doc*).

Once the 145 files are extracted from the *ca_all_2006.zip file*, header information for the geographic file is obtained from the technical documentation (Table 2, page 12). Then, the data dictionary is read into SPSS and used to generate the base code needed to read in the approximately 27,000 variables from the 145 files. The base code then requires editing so that it can read in all of the variables from the 145 files and merge them into a final data file containing all of the ACS data for all census geographies available for that year. This process is similar for all the ACS data files for every year, as well as for those ACS years in which the datasets include 3-year and 5-year averages.

The Census 2000 and 2010 data file collection process and standardization procedure is similar to the ACS procedures. The exception is that Census 2000 data files need to be collected and processed separately for Summary File 1 (SF1) and Summary File 3 (SF3). SF1 contains population-related variable and SF3 contains housing-related variables. Census 2010 only produced SF1 data; for 2010 and beyond, the equivalent to the Census 2000 SF3 variables are provided in ACS data files, but only as 5-year averages. For example, SF3 variables for 2008 are provided in ACS 2010, and SF3 variables for 2009 will be provided in ACS 2011, to be released in November, 2012.

Having completed all of the data file downloading and initial processing, *Demand Research* now has on hand all of the Census and ACS data files listed in Table 2, both in their original forms, and more importantly, in standardized datasets. This allows for quick access to all of the data and to individual variables, and for data verification when questions arise. Note that the standardized datasets are of considerable size and thus no further manipulation is done with them. Rather, variables from the standardized datasets are selected, by the desired geographic level, and imported into new, smaller files for further processing.

Weather Data

Local weather data from the National Climatic Data Center (NCDC), that is, monthly heating degree days and cooling degree days (HDD and CDD, base 65 degree Fahrenheit), require special processing. NCDC provides data file S3220 for all cooperative weather stations in California for the years 2004 through 2010. The number of weather stations vary from year to year (some were retired and others came online), but in no year are there less than 300 weather stations.

HDD and CDD were inspected for missing monthly data and those weather stations for which there were two or more consecutive months of missing data were dropped. This resulted in a final sample of 153 weather stations that end up being located in 51 of the 58 California counties. Where a single month of HDD or CDD was missing, its value was interpolated from the average of the two adjacent months.

Based on information provided by the California Energy Commission (CEC), the state is separated into 16 climate zones that are county and utility-specific and that are associated with unique U.S. postal service zip codes. To assign weather stations to local areas, the state zip codes were first matched to Census 2010 zip code analogs, Zip Code Tabulation Areas, known as ZCTAs or ZCTA5s for 5-digit zip codes. Of California's 710,145 census blocks, 90 percent were matched from ZCTA5 to zip codes. The remaining 67,388 census blocks were linked to their closest climate zone and weather station via census block groups and tracts. After initial local assignment, the distance between weather stations and census blocks were calculated. In cases where more than one weather station was associated with a given climate zone, census blocks were linked to the closest weather station.

In all, every climate zones except climate zone 16 contained one or more weather stations. As climate zone 16 represents a small portion of Pasadena and Glenwood in Los Angeles County, its census tracts were linked to adjacent climate zone 9, which also represent portions of Pasadena.

Additional Data, as Needed

The remaining datasets listed in Table A1 are relatively small compared to the Census and ACS datasets. These have been downloaded from the various agencies and organization as either Excel of CVS files, and have been stored in SPSS files with formats that conform to the Census and ACS files so that their variables are easily identified and imported into analysis files.

A3. Census Tract Variables

This section describes the procedures used for creating continuous, consistent population and housing characteristics variables for all census tracts in California between the years 2006 and 2010. Because of the different features of Census 2000 and Census 2010 and ACS 2006 through ACS 2010, several types of technical procedures are employed to create the needed demographic and economic variables. So far as can be discerned from a literature search, there are no published papers or reports that describe a methodology for undertaking tasks of this kind, perhaps because the data needed for producing reliable intercensal census tract-level variables have only become available within the past 14 months.

Demand Research is employing several different mathematical formulas for imputing intercensal census tract variables for 2006 through 2010. *Imputation* is a generic term that refers to substituting a value for a missing data point. Specific imputation procedures used for the Census and ACS datasets are:

- *interpolation and extrapolation:* this involves constructing new data points within the range of a discrete set of known beginning and end data points, or constructing new data points when the end point is unknown.

- *geospatial standardization and allocation:* because census tract boundaries changed, and new tracts were added, from the 2000 to the 2010 Census, data collected for the years prior to 2010 must be adjusted so that they conform to the new boundaries. *Allocation* involves attributing fractions of a value to different subjects (e.g. census tracts) based on an explicit weighting scheme.
- *reconciliation:* this involves adjusting values that are derived by interpolation, standardization and allocation so that their sum equals a known value, for example, so that the sum of all imputed census tract populations within a county equals the county's known population.

It is important to note that the order in which these procedures are performed is critical to the success of the imputations – and the procedure order differs for different years. This is entirely due to the fact that there are different relevant datasets for each year, and the intention of this pilot project is not to minimize and standardize data processing procedures from year to year, but rather to use the maximum amount of data for minimizing census tract-level measurement error.

Interpolation

The most common form of time series interpolation involves constructing a constant value that can be used to bridge the known beginning and end values of a data series. One such constant value is an annual average growth rate (AAGR). For a given census tract, it is defined as:

$$AAGR = (Value_t \,/\, Value_{t-n})^{(1/(t-(t-n)))}$$

where $Value_t$ is the end point value at time t of the variable and $Value_{t-n}$ is the beginning point value of the variable at some year prior to t. Alternatively, an annual average growth amount (AAGA) can be constructed as:

$$AAGA = (Value_t - Value_{t-n}) \,/\, (t-(t-n))$$

The choice one or the other function for interpolation depends on the assumptions made about the likely growth pattern of the variable in question. However, it is worth noting that when the known beginning and ending values are within about 3 percent of each other, as is the case for many variables in the intercensal years, AAGR and AAGA constants produce almost identical interpolated values.

Another form of interpolation, year-over-year percent change (YOY), produces different interpolation factors for each year. This function can be used when annual data are available for one geographic area but not another, and the area without data has a beginning value. It assumes that the area without complete data has values that change in exactly the same proportion as the values of the area with complete data. For example, if annual population counts, *C*, are available for all years for county *i*, but not for any of the *j* census tracts within county *i*, then census tract population counts can be imputed by multiplying their known beginning year values by the county YOY values, defined as:

$$YOY_{i,t} = (C_{i,t} / C_{i,t-1})$$

The imputed values for the *j* census tracts are then calculated as:

$$C_{j,t} = C_{j,t-1} \times YOY_{i,t}$$

which naturally sum to the county population counts in each year.

Extrapolation can take many forms, depending on the variable. The difference between interpolation and extrapolation is that extrapolation is applied when there is no sufficient data to tie a missing value between known end points. Extrapolations can be done using AAGR or AAGA constants to extend the data into future years. Or various other forms of trending and forecasting, both simple or multivariate can be employed.

Geospatial Standardization and Allocation

The goal of the pilot project is to perform multivariate statistical analyses of energy use using panel data, that is, data spanning multiple years and multiple geographic areas. To accomplish this goal it is essential that all the years of data represent common census tract geographies. This does not present a problem if the boundaries of the geographic areas in question never change, such as with counties and states, but does present a problem with census tracts, many of whose boundaries change for political and demographic reasons from decade to decade.

As can be seen in TableA2, Census 2000 provides census tract-level data using the 7,049 census tracts and boundaries, while Census 2010 and ACS 2010 offer census tract-level data for 8,057 census tract. Based on the geospatial crosswalk factors provided by the Census Bureau in their census tract relationship file, many census tract boundaries, as defined by polygons, remained the same from 2000 to 2010. However, in addition to more than 1,000 new census tracts, many of the census tract boundaries changed.

Table A3 contains a list of several of the variables in the Census Bureau census tract relationship file.

Table A3. Census 2010 Geospatial Standardization Factors

Variable[1]	**Description**
HU10PT	Calculated 2010 Housing Unit count for the relationship record
POP10PT	Calculated 2010 Population for the relationship record
COUNTY00	2000 County Code
COUNTY10	2010 County Code
TRACT00	2000 Census Tract Code
TRACT10	2010 Tract Code
HU00	2010 Housing Unit Count of the 2000 Tract
HU10	2010 Housing Unit Count of the 2010 Tract
POP00	2010 Population of the 2000 Tract
POP10	2010 Population of the 2010 Tract
HUPCT00	Calculated Percentage of the HU00 this record (HU10PT) contains
HUPCT10	Calculated Percentage of the HU10 this record (HU10PT) contains
POPPCT00	Calculated Percentage of the POP00 this record (POP10PT) contains
POPPCT10	Calculated Percentage of the POP10 this record (POP10PT) contains

[1]From the Census Bureau, 2012. All variables in this table are produced , published, and documented by the Census Bureau and can be found on their website

Other than two census tract area variables, adjustment factors are provided based on two key variables, i.e., census tract population and census tract housing units. These factors are explicitly created for standardizing Census 2000 census tract data to the Census 2010 census tracts (see the PDF file "Understanding the 2010 Census Tract Relationship Files" for examples of how these factors are to be used). Thus, the same adjustment factors are used for all ACS and imputed data for the years prior to the new boundaries in 2010. Of course, when a 2010 census tract has the same boundaries it had in 2000, no adjustments is necessary. But when a census boundary is merged or split or otherwise revised, one or more procedures are needed to standardize the data for 2010 by allocating values according to the 2010 change. Table 3 contains the Census Bureau's mnemonics and descriptions for population and housing unit variables in the census tract relationship file.

The actual census tract adjustment variables are those in Table A3 ending in *PCT00* and *PCT10*. *POPPCT00* is the percent of the 2010 population contained in that portion of the 2000 census tract that is also contained in the 2010 Census tract of that record, and *POPPCT10* is the percent of the population of the 2010 census tract that this record contains. The same goes for the housing unit count adjustment factors.

For variables measured in counts, such total employment or number of elderly persons, the *PCT00* variables must be applied. However, the *PCT00* adjustment factors cannot be used for adjusting variables that are measured as averages or percents, such as average income or percent housing vacancies. For these variables the *PCT10* adjustment factors are used as weighting factors.

Reconciliation

Reconciliation is the final imputation procedure. Once census tract variable values are imputed for each year, each year's sum of the census tract values must be equal to the known value of the county in which the tracts are located. To ensure that this occurs, each census tract value is upgraded or downgraded by the percentage difference between the sum of the census tract values and the known county control total. As a first pass at reconciliation, the one-year estimates will be used. However, these only cover 40 of the 58 counties for 2006 through 2009 (and SF3 variables for 2010). As a result, the 3-year averages will be introduced for the additional 11 counties for 2006 and 2009. Reconciliation is not possible for the data for the final 7 counties for these years, nor for the SF3 variables for 2010 for 18 counties.

Census Tract Data Development Procedures for 2006, 2007, 2008, 2009, and 2010

To explain how these procedures are applied, Table A4 lists the available Census and ACS datasets by year, as well as geographical level. This not only shows where the gaps are in the census tract population and housing data, but which datasets are available for making the necessary intercensal census tract imputations.

Table A4. Geo-temporal Coverage of California Census and ACS Data

Year Represented	1 Year Tracts	1 Year Counties	3 Year Average Tracts	3 Year Average Counties	5 Year Average Tracts	5 Year Average Counties	Data Sources
2000	7,049	58	--	--	--	--	Census 2000
2001	--	--	--	--	--	--	--
2002	--	--	--	--	--	--	--
2003	--	--	--	--	--	--	--
2004	--	--	--	--	--	--	--
2005	--	40	--	--	--	--	ACS 05
2006	--	40	--	51	--	--	ACS 06, 07
2007	--	40	--	51	7,049	58	ACS 07, 08, 09
2008	--	40	--	51	8,057	58	ACS 08, 09, 10
2009	--	40	--	51	--	--	ACS 09, 10
2010		40 (sf3)					ACS 10
2010	8,057 (sf1)	58 (sf1)	--	--			Census 2010

One important consideration is that although actual census tract data are available for 3 of the 5 study year, the data taken to represent 2007 and 2008 are ACS 5-year averages, not single year data. Although there is no census tract data with which to verify that this designation is a good approximation of the 1-year conditions, comparisons of county-level 5-year average estimates with single year estimates for the mid-year average show that these two values are in very close agreement. This is what can reasonably be expected, because economic, demographic, and housing conditions change slowly.

In Table A5, BEA's single-year local area personal income and employment data for 2007 and 2008 are compared with the ACS county-level 5-year averages. As can be seen, when aggregating the 58 county data to the state level, the two population estimates are almost identical for each year, and the two employment estimates tend to differ by similar magnitudes in each year, with the ACS statistics being 11.5 and 13.7 percent lower than the BEA statistics. Further inspection of the employment data shows that the county-level differences between the ACS and BEA data tend to be consistent from year to year. This suggests that the employment differences are most likely related to the different ways in which employment is defined and measured. ACS employment statistics are constructed using personal responses to survey questions, while the BEA employment statistics are created from a complex combination of survey responses, census data, Internal Revenue Service data, and other administrative agency data. In all, the correlations for both variables in both years round to unity, indicating that the differences between the two versions of the data in each year are highly systematic.

Table A5. California County-Level Data, 2007 and 2008

Year: 2007	**Population**	**Employment**
ACS 5-Yr. Average	36,308,527	18,100,948
BEA 1-Yr. Estimate	36,226,122	20,965,534
% Difference	0.23%	-13.66%
Avg. % Diff. (n=58)	-0.03%	-7.32%
Correlation (n=58)	1.00	1.00
Year: 2008		
ACS 5-Yr. Average	36,637,290	18,418,306
BEA 1-Yr. Estimate	36,580,371	20,810,649
% Difference	0.16%	-11.50%
Avg. % Diff. (n=58)	1.01%	-4.56%
Correlation (n=58)	1.00	1.00

Another further test of how well the 5-year averages represent the mid-year of the average compares the county-level 1-year Census data with the 5-year averages for the census tracts when aggregated to the county-level. As can be seen from Table 4, this can be done for 2007 and 2008 for the 40 counties for which there are 1-year estimates. Table 6 provides the same descriptive statistics as Table 5 for population and employment, as well as for the number of electrically-heated homes in the counties, the median household income, and the median age of the population. As is readily apparent, the 5-year averages are very close matches to the 1-year estimates in both years for all five variables.

Assigning the ACS 5-year estimates to 2007 and 2008 greatly simplifies the task of creating a complete census tract-level dataset, as does the availability of 1-year estimates for all census tracts for all SF1 variables in 2010 from the 2010 Census. However, there remains the need to perform imputations for the SF1 and SF3 variables for 2006 and 2009, and the SF3 variables 2010. Table 7 contains descriptions of the basic procedures followed for constructing the census tract datasets for each year. Note that these imputation procedures are not necessarily final, and that *Demand Research* anticipates that once these procedures are reviewed by the project advisory panel, or other technical experts, they may be improved. Since the all of the datasets are now in place in SPSS and readily available for processing, changes to these procedures will not require a great deal of extra effort.

Table A6. Comparison of County-Level 1-Year and 5-Year Estimates for 2007 and 2008

Year: 2007	Population	Employment	Elect.-Heat Homes[1]	Median HH Inc.	Median Age
ACS 5-Yr. Average	35,800,118	17,791,441	2,789,540	$60,875	35.9
ACS 1-Yr. Estimate	36,041,596	17,823,902	2,808,087	$56,994	35.2
% Difference	-0.67%	-0.18%	-0.66%	6.81%	1.87%
Avg. % Diff. (n=40)	-0.57%	-2.77%	-0.59%	6.85%	1.97%
Correlation (n=40)	1.00	1.00	1.00	0.98	0.98
Year: 2008					
ACS 5-Yr. Average	36,123,617	18,108,163	2,859,007	$61,660	36.5
ACS 1-Yr. Estimate	36,246,049	18,279,230	2,773,119	$58,925	35.3
% Difference	-0.34%	-0.94%	3.10%	4.64%	3.38%
Avg. % Diff. (n=40)	0.23%	-3.06%	2.57%	4.94%	3.53%
Correlation (n=40)	1.00	1.00	1.00	0.98	0.95

[1]due to missing values for 2007 (n=39) and for 2008 (n=36)

Procedures for Imputing Missing Values for 2009 and 2010 Census Variables

Understanding how demographic and economic conditions changed over the 5-year period from 2006 to 2010 in California is essential for choosing the best imputation procedures for the 2009 SF1 and SF3 variables, and the 2010 SF3 variables. To gain insight into these conditions, state-level trends are analyzed.

Graph A1 shows California's population trend for 2006 through 2010. This trend has a steady growth rate of approximately 1 percent annually, as can be seen by the year-over-year percentage changes presented in Table A7.

Graph A1: State of California Population, 2006-2010
(Sources: BEA)

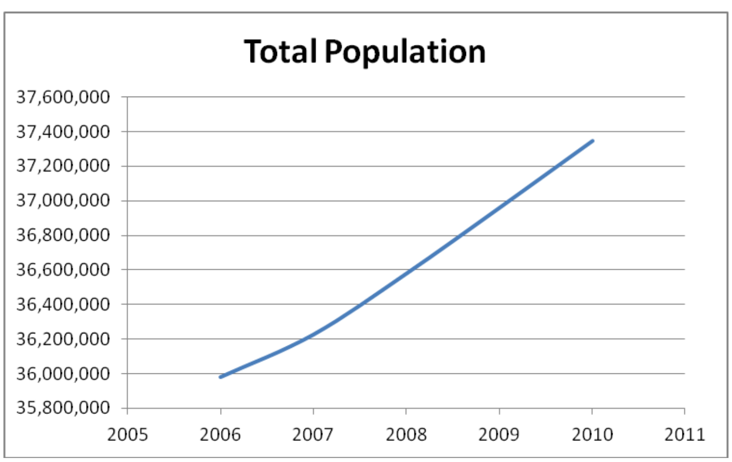


Table A7: Year-Over-Year Percentage Change in Population

Year	% Change
2006	
2007	1.007%
2008	1.010%
2009	1.010%
2010	1.010%

Given the nearly constant proportional population trend at the state level, the Census variables related to population are imputed by applying a steady, one-directional trend to missing values. Where 2008 and 20010 data are available for a variable, *Y*, in census tract *i*, that is measured in frequencies or counts, but a value is not available for 2009, the midpoint will be assigned to 2009, calculated as:

$$\text{Imputed } Y_{i,2009} = \text{Midpoint } Y_{i,\ 2008\text{-}2010} = \frac{\left(Y_{i,2010} - Y_{i2008}\right)}{2} + Y_{i,2008}$$

Where a count variable, *W*, is missing both 2009 and 2010 values, the census-level growth rate for a known variable or proxy, such as population or housing units, will be applied to it to impute the 2009 and 2010 values. In this case, the percentage change for the midpoint of the proxy variable must first be calculated. Taking population *(Pop)* as the proxy:

$$\%\text{ Change in } Pop_{i,\ 2008\text{-}2010} = \%\Delta Pop_i = 1 + \left(\frac{\left(Pop_{i,2010} - Pop_{i,2008}\right)}{Pop_{i,2008}}\right) / 2$$

The percentage changes for *POP* is then be multiplied by the 2008 values of *W* to get the imputed 2009 and by the imputed value for 2009 to get the 2010 value for *W*.

$$\text{Imputed } W_{i,2009} = W_{i,2008} \times \%\Delta Pop_i$$
$$\text{Imputed } W_{i,2010} = \text{Imputed } W_{i,2009} \times \%\Delta Pop_i$$

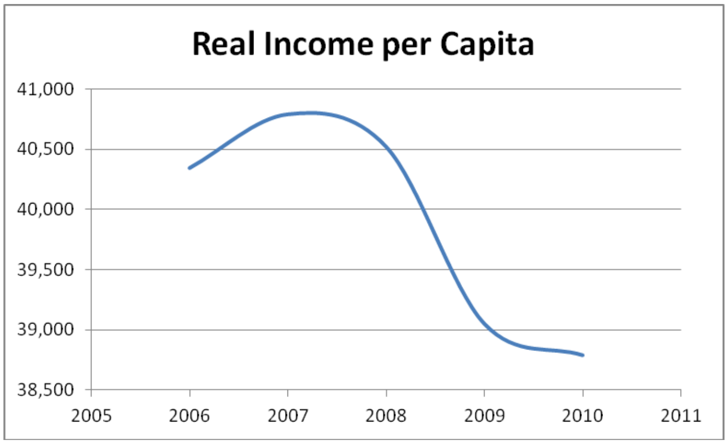
Not all Census variables are counts; many are medians or means, and others are age or category-related. Specific, if simpler, routines are needed to impute the values for these variables. Table A8 contains the variables on which this procedure is performed.

Table A8: Variables in Situation 1

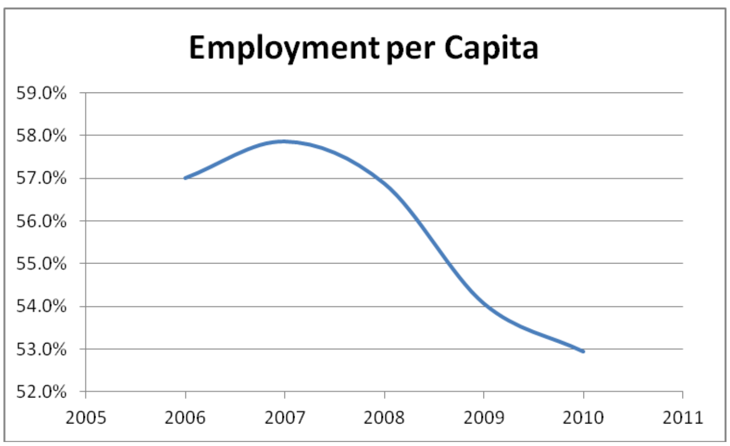
Variable	Data Source	Missing	Procedure	Other Uses
Population	Census 2010 SF1 - 1 yr	2009	simple	this is a used as a weight
HHold	estimate for 2010 and	2009	interpolation	this is used as a weight
HHoldpop	ACS 2010 SF1 - 5yr	2009	of midpoint	
HHunder18	estimate for 2008	2009	for	
HHover59		2009	2009	
HHover64		2009		
Grouppop		2009		
HUtotal		2009		this is used as a weight
OccupiedHU		2009		
VacantHU		2009		
OwnerHU		2009		this is used as a weight
RenterHU		2009		
AvgsizeHH		2009		this is used as a weight

The imputation procedures used for variables that generally exhibit monotonic trends are not appropriate for variables that whose trends turned direction. Graphs A2 and A3 show reversals in trends for two key indicators of economic conditions, i.e., real per capita income and employment per capita. These suggest that the imputation procedure used for economic variables should attempt to capture the effects of the downturn. These variable hit their peaks in 2007 and then deteriorated in 2008, 2009, and 2010.

Graph A2: State of California Real Income per Capita (2005$), 2006-2010
(Source: BEA)



Graph A3: State of California, Percent of Population Employed, 2006-2010
(Source: BEA)



For the Census variables directly related to the business cycle, such as total personal income and total employment, the imputation procedures depend on which year of data is missing and what data are available in the different years and at the different geographic levels. To combine as much good information as possible and to avoid reliance on a single year-over-year (YOY) change that might be anomalous, growth rates are nested and averaged for these variables. Given an economic variable, *Z*, the following assignments are applied to geographic areas:

- let *(a)* refer to census tracts in the 40 largest counties (Pop >65,000)

- let *(b)* refer to census tracts in the next largest 11 counties (Pop >20,000 <65,000)
- let (*c)* refer to census tract in the smallest 7 counties (Pop <20,000) (note: that there are no SF3 data for the census tracts in these counties for 2009 or 2010) such that:

$$a + b + c = i = 8{,}057 \text{ census tracts}$$

- let *(j)* refer to each of the 40 counties (Pop >20,000)
- let *(k)* refer to each of the 11 counties (Pop >65,000)
- let *(l)* refer to each of the 7 smallest counties (Pop <20,000)
- let *(m)* refer to the 51 largest counties (Pop>20,000) such that:

$$j + k + l = n = 58 \text{ counties, and}$$

$$j + k = m = 51 \text{ counties}$$

a) <u>For the census tracts *(a)* – these tracts have county-level and tract level data for 2008, 2009, 2010:</u>

Step 1: Calculate the YOY 2007-2008 change for (*a)* census tracts using the 2007 and 2008 5-year ACS averages:

$$YOY_{a,2007-2008} = Z_{a,2008} / Z_{a,2007}$$

Step 2: Calculate the YOY 2008-2009 change for the 40 *(j)* counties using the 2008 and 2009 1-year ACS averages:

$$YOY_{j,2008-2009} = Z_{j,2009} / Z_{j,2008}$$

Step 3: Calculate the YOY 2010-2009 change for all the 40 *(k)* counties using the 2009 and 2010 1-year ACS averages:

$$YOY_{j,2009-2010} = Z_{j,2010} / Z_{j,2009}$$

Step 4: Average the 3 nested YOY for each of the *(a)* census tracts. This produces a percentage change that is to be applied each *(a)* census tract in 2009, and then 2010:

$$\text{Imputed } Z_{a,2009} = \frac{(YOY_{j,2010-2009} + YOY_{j,2009-2008} + YOY_{a,2008-2007})}{3} \times Z_{a,2008}$$

$$\text{Imputed } Z_{a,2010} = \frac{(YOY_{j,2010-2009} + YOY_{j,2009-2008} + YOY_{a,2008-2007})}{3} \times \text{Imputed } Z_{a,2009}$$

b) <u>For the census tracts *(b)* – these tract have county-level and tract-level data for 2008 and 2009*:*</u>

For the census tracts *(b)* that are in the 11 counties for which there are no county-level data for 2010, but for whom there are 2009 county-level data

Step 1: Calculate the YOY 2007-2008 change for (*b)* census tracts using the 2007 and 2008 5-year ACS averages:

$$YOY_{b,2007-2008} = Z_{b,2008} / Z_{b,2007}$$

Step 2: Calculate the YOY 2008-2009 change for the 11 *(k)* counties using the 2008 and 2009 3-year ACS averages:

$$YOY_{k,2008-2009} = Z_{k,2009} / Z_{k,2008}$$

Step 3: Calculate the mean YOY 2010-2009 (MYOY) change for all the 40 *(j)* counties using the 2009 and 2010 1-year ACS averages:

$$MYOY_{j,2009-2010} = \frac{\sum Z_{j,2010}}{\sum Z_{j,2009}}$$

Step 5: Average the 3 nested growth rates and apply to *(b)* census tracts:

$$\text{Imputed } Z_{b,2009} = \frac{(MYOY_{j,2010-2009} + YOY_{k,2009-2008} + YOY_{b,2008-2007})}{3} \times Z_{b,2008}$$
$$\text{Imputed } Z_{b,2010} = \frac{(MYOY_{j,2010-2009} + YOY_{k,2009-2008} + YOY_{b,2008-2007})}{3} \times \text{Imputed } Z_{b,2009}$$

c) <u>For the census tracts *(c)* – these tracts have county-level and tract level data for 2008, only:</u>

Step 1: Calculate the YOY 2007-2008 change for *(c)* census tracts using the 2007 and 2008 5-year ACS averages:

$$YOY_{c,2007-2008} = Z_{c,2008} / Z_{c,2007}$$

Step 2: Calculate the YOY 2008-2009 change for the 51 *(m)* counties using the 2008 and 2009 3-year ACS averages:

$$YOY_{m,2008-2009} = Z_{m,2009} / Z_{m,2008}$$

Step 3: Calculate MYOY for *(m)* counties for 2008-2009 as:

$$MYOY_{m,2008-2009} = \frac{\sum Z_{m,2009}}{\sum Z_{m,2008}}$$

Step 5: Average the 3 nested growth rates and apply to *(c)* census tracts:

$$\text{Imputed } Z_{c,2009} = \frac{(MYOY_{j,2010-2009} + MYOY_{m,2009-2008} + YOY_{c,2008-2007})}{3} \times Z_{c,2008}$$
$$\text{Imputed } Z_{c,2010} = \frac{(MYOY_{j,2010-2009} + MYOY_{m,2009-2008} + YOY_{c,2008-2007})}{3} \times \text{Imputed } Z_{c,2009}$$

d) For the census tracts with suspiciously large percentage changes from 2007 to 2008 (e.g., > +/ -0.25).

DEFAULT RATE IF $YOY_{i,2008\text{-}2007} > 0.50$ in Census Tracts (a):

$$\text{Imputed } Z_{a,2009} = \frac{(YOY_{j,2010-2009} + YOY_{m,2009-2008} + YOY_{n,2008-2007})}{3} \times Z_{a,2008}$$
$$\text{Imputed } Z_{a,2010} = \frac{(YOY_{j,2010-2009} + YOY_{m,2009-2008} + YOY_{n,2008-2007})}{3} \times \text{Imputed } Z_{a,2009}$$

DEFAULT RATE IF $YOY_{i,2008\text{-}2007} > 0.50$ in Census Tracts (b):

$$\text{Imputed } Z_{b,2009} = \frac{(MYOY_{j,2010-2009} + YOY_{m,2009-2008} + YOY_{n,2008-2007})}{3} \times Z_{b,2008}$$
$$\text{Imputed } Z_{b,2010} = \frac{(MYOY_{j,2010-2009} + YOY_{m,2009-2008} + YOY_{n,2008-2007})}{3} \times \text{Imputed } Z_{b,2009}$$

DEFAULT RATE IF $YOY_{i,2008\text{-}2007} > 0.50$ in Census Tracts (c):

$$\text{Imputed } Z_{c,2009} = \frac{(MYOY_{j,2010-2009} + MYOY_{m,2009-2008} + YOY_{n,2008-2007})}{3} \times Z_{c,2008}$$
$$\text{Imputed } Z_{c,2010} = \frac{(MYOY_{j,2010-2009} + MYOY_{m,2009-2008} + YOY_{n,2008-2007})}{3} \times \text{Imputed } Z_{c,2009}$$

Finally, note that because of the way some variables are measured, they are not amenable to any of these imputation approaches. For example, values for the change in the median age of the housing stock cannot be imputed using any of these approaches. For variables that do not fit into a pre-defined procedure, imputations will be made on a case-by-case basis. In most such cases, the best estimate is often derived from a simple rule that follows from the logic of the variable.

APPENDIX B: Annual Energy Consumption and Program Tracking Data

At the inception of this pilot project *Demand Research* requested from the Commission's Energy Division (ED) annual electric and gas utility billing data and demand-side management program tracking data. At this time, due to prior ED requests, investor-owned utilities were providing residential billing and program tracking data to KEMA, and non-residential billing and program tracking data to ITRON. These data were for calendar year 2006 through calendar year 2010.

Demand Research, working in coordination with KEMA and ITRON, received all five years worth of residential electricity customers and residential natural gas customers data for SDG&E and PG&E, but not for SCE . *Demand Research* also received non-residential electricity customer data for SDG&E, PG&E, and SCE, and nonresidential natural gas customer data for SDG&E and PG&E, but not for SGC. The key variables that *Demand Research* received were:

e) total annual retail fuel (electricity or natural gas) sales, by year and census tract
f) total annual number of premises served, by fuel, year, and census tract
g) total annual ex ante gross energy savings, by fuel, year, and census tract
h) total annual energy efficiency measure costs by fuel, year, and census tract
i) for nonresidential customers only, total incentives paid to customers, by fuel, year, and census tract.

Due to technical difficulties, the residential data could no be disaggregated by type of premise (single family, multifamily, etc.). However, the residential data were disaggregated by 2 and 3 digit NAICS codes.

Because utility databases do not identify the census tract for each account or premise, both KEMA and ITRON developed "address normalization" routines using either their own procedures or purchased software to assign census tracts to each customers. This effort was largely successful; however, a small fraction of customer premises remained unassigned or were assigned incorrect census tracts.

In addition to address normalization, KEMA and ITRON were asked to produce hypothetical census-level energy expenditures. KEMA described the way they did so as follows:

> For each month, we split the total consumption into as many tiers as there are under the prevailing rate. For example, if the bill is 439 kWh and there are 4 tiers:
>
> - The first 200 kWh go on the first tier (baseline – the lowest price)
> - Next 200 kWh go on the second tier (higher price)
> - Last 39 kWh go on the third tier (even higher price)
> - This customer doesn't have any kWh in the fourth tier (highest price) this month.
>
> The kWh per month allowed in the first tier vary with the geographical location of the customer (for example, PG&E has about 10 baseline territories) and the season. We took this into account.
> We worked with prevailing rates (number of tiers and price per kWh) for each year.
> We added up the monthly dollars to come up with annual dollar amounts at the customer level.
> Then we added up the annual dollars for each customer in each census tract.
>
> (*email from Paula Ham-Su, KEMA, to M. Horowitz, May 22, 2012*)

To check on their estimates, KEMA compared their total expenditures in the residential sector to those reported by the utilities to EIA and found that they closely matched. ITRON's approach was similar to KEMA's approach. ITRON described its methodology as follows:

> In short, we modeled the rates for 2009 for all the most common (and largest consuming) rate codes. We applied these rates at the account level for each month, summed the months to a year, summed the accounts to a site, and then averaged (sic, it should have been "totaled") the bills at the census tract/naics. There were a lot of rate codes and some of them only accounted for a tiny part of the IOU's consumption (in total, about 2% of consumption and 5% of accounts). The accounts on those rates were left out of the calculation.
>
> *(email from Christine Hungeling, ITRON, to M. Horowitz, April 13, 2012)*

APPENDIX C: MCM Pilot Study Data Dictionary

Tables C1 and C2 contain listings of most of the variables in the analysis datasets along with a brief description of the variables and their sources. Table C1 is for the residential sector and Table C2 is for the nonresidential sector. The listed variables in no particular order. They are subsets of the thousands of variables that were downloaded from government databases and websites and processed for use. They were initially chosen based on their expected contributions to the modeling effort. These variables are not only a subset of all the variables collected, but a subset of all the variables that were created in the course of implementing the statistical analyses.

Note that all of the variables in the tables were worked on at some time, but not all of these variables found their way into the final models presented in the study report. Also, it is important to note that there are more IOU billing and program tracking variables available than those listed in these tables. The variables listed are those provided by KEMA and ITRON that ended up being used in the analyses.

Table C1: Residential Sector Variables (Utility Data Provided by KEMA)

Variable Name	Variable Label	Source
county	County FIPS code	KEMA
tract	Census tract	KEMA
year	Year	Utility
utility	Utility code	Utility
elec	Number of electric only accounts	Utility
gas	Number of gas only accounts	Utility
both	Number of electric and gas accounts	Utility
e_usage	kWh total annual consumption	Utility
g_usage	Therms total annual consumption	Utility
e_tot_bill	kWh total annual bill	KEMA
g_tot_bill	Therms total annual bill	KEMA
IOUexanteGrSavkW	kW annual savings	Utility
IOUexanteGrSavkWh	kWh annual savings	Utility
IOUexanteGrSavTherms	Therms annual savings	Utility
IOUGrMeaCost	Total gas electric measure cost	Utility
IOUGrMeaCost_e	Total electric measure cost	Utility
IOUGrMeaCost_g	Total gas measure cost	Utility
IOUexante_kW_count	Number of KW ECMs	Utility
IOUexante_kWh_count	Number of kWh ECMs	Utility
IOUexante_Thm_count	Number of gas ECMs	Utility
nmeas_any_1up	N of accounts with 1 or more gas or electric ECM	Utility
nmeas_elec_1	Number of accounts with one electric ECM	Utility

Variable Name	Variable Label	Source
nmeas_elec_2	Number of accounts with two electric ECMs	Utility
nmeas_elec_3	Number of accounts with three electric ECMs	Utility
nmeas_elec_4up	Number of accounts with>=4 electric ECMs	Utility
nmeas_gas_1	Number of accounts with one gas ECM	Utility
nmeas_gas_2	Number of accounts with two gas ECMs	Utility
nmeas_gas_3	Number of accounts with three gas ECMs	Utility
nmeas_gas_4up	Number of accounts with>=4 gas ECMs	Utility
lowincome	Number of low income accounts	Utility
population	Population	Census
medianage	Median age	Census
hhold	Number of households	Census
hholdpop	Household population	Census
hhunder18	Household with someone under 18	Census
hhover59	Household with someone over 59	Census
hhover64	Household with someone over 64	Census
grouppop	Population in group living situation	Census
pop25andover	Population 25 years of age or older	Census
noschooling	No schooling	Census
nurseryto4thgrade	Nursery to 4th grade	Census
grade5thand6th	Grade 5th and 6th	Census
grade7thand8th	Grade7th and 8th	Census
grade9th	Grade 9th	Census
grade10th	Grade 10th	Census
grade11th	Grade 11th	Census
grade12thnodiploma	Grade12th no diploma	Census
highschoolgraduate	High school graduate	Census
collegelessthan1year	College less than 1 year	Census
college1plusyrsnodegree	College 1or more years nodegree	Census
associatedegree	Associate degree	Census
bachelorsdegree	Bachelors degree	Census
mastersdegree	Masters degree	Census
professional	Professional degree	Census
doctorate	Doctorate	Census
utilitygas	Utility gas	Census
bottledorlpgas	Bottled or LP gas	Census
electricity	Electricity	Census
fueloilkeroseneetc	Fuel oil, kerosene, etc.	Census
coalcoke	Coal or coke	Census
wood	Wood	Census
solarenergy	Solar energy	Census
otherfuel	Other fuel	Census
nofuel	No fuel	Census
built2005orlater	Built2005 or later	Census
built2000to2004	Built 2000 to 2004	Census
built1990to1999	Built 1990 to 1999	Census
built1980to1989	Built 1980 to 1989	Census
built1970to1979	Built 1970 to 1979	Census
built1960to1969	Built 1960 to 1969	Census
built1950to1959	Built 1950 to 1959	Census

Variable Name	Variable Label	Source
built1940to1949	Built 1940 to 1949	Census
built1939orearlier	Built 1939 or earlier	Census
medianyearbuilt	Median Year built	Census
medianbuiltyears	Median age of building	Census
hu1detached	Single family detached	Census
hu1attached	Single family attached	Census
hu2	Duplex	Census
hu3or4	Triplex/Quad	Census
hu5to9	building 5-9 units	Census
hu10to19	building 10-19 units	Census
hu20to49	building 20-49 units	Census
hu50ormore	building 50 or more units	Census
humobilehome	Mobile home	Census
huboatrvvan	Boat, RV or van	Census
room1	Housing units with 1 room	Census
room2	Housing units with 2 rooms	Census
room3	Housing units with 3 rooms	Census
room4	Housing units with 4 rooms	Census
room5	Housing units with 5 rooms	Census
room6	Housing units with 6 rooms	Census
room7ge	Housing units with 7 or more rooms	Census
medianrooms	Median number of rooms	Census
hutotal	Total number of housing units	Census
occupiedhu	Occupied housing units	Census
vacanthu	Vacant housing units	Census
ownerhu	Owner occupied housing units	Census
renterhu	Renter occupied housing units	Census
inclt10	Households with <$10,000 annual income	Census
inc10to15	Households with $10,000-$15,000 annual income	Census
inc15to20	Households with $15,000-$19,999 annual income	Census
inc2025	Households with $20,000-$24,999 annual income	Census
inc25to29	Households with $25,000-$29,999 annual income	Census
inc30to34	Households with $30,000-$34,999 annual income	Census
inc35to39	Households with $35,000-$39,999 annual income	Census
inc40to44	Households with $40,000-$44,999 annual income	Census
inc45to49	Households with $45,000-$49,999 annual income	Census
inc50to59	Households with $50,000-$59,999 annual income	Census
inc60to74	Households with $60,000-$74,999 annual income	Census
inc75to99	Households with $75,000-$99,999 annual income	Census
inc100to124	Households with $100,000-$124,999 annual income	Census
inc125to149	Households with $125,000-$149,999 annual income	Census
inc150to199	Households with $150,000-$199,999 annual income	Census
incgt200	Households with $200,000 or more annual income	Census
medianincome	Median household income	Census
addhhincome	Aggregate household income	Census
personhh1	1 person household	Census
personhh2	2 person household	Census
personhh3	3 person household	Census
personhh4	4 person household	Census

Variable Name	Variable Label	Source
personhh5	5 person household	Census
personhh6	6 person household	Census
personhh7	7 or greater person household	Census
avgsizehh	Average household size	Census
avghhsizeowner	Average household size Owner occupied housing units	Census
avghhsizerenter	Average household size renter occupied housing units	Census
aggssinchh	Aggregate household income from Social Security	Census
hhwithssincome	Households with Social Security Income	Census
hhwithnossincome	Households with no Social Security Income	Census
pop16andover	Population over 16 years of age or older	Census
Inlaborforce	Population in labor force	Census
inarmedforces	Population inarmed forces	Census
civilian	Civilian population	Census
employed	Employed	Census
unemployed	Unemployed	Census
notinlaborforce	Not in labor force	Census
povertyhh	Housholds in poverty	Census
nonpovertyhh	Households not in poverty	Census
aggrentasked	Aggregate rent asked	Census
avgrentasked	Average rent asked	Census
mediancontractrent	Median contract rent	Census
mediangrossrent	Median gross rent	Census
medianhuvalue	Median value of housing unit	Census
huwithmortgage	Housing units with mortgage	Census
hu2ndmortgage	Housing units with second mortgage	Census
huhomeequityloan	Housing units with home equity loan	Census
hu2ndmortandhel	HU with second mortgage & home equity loan	Census
humortgagenoheloan	Housing units with no home equity loan	Census
huwithno2ndmortgage	Housing units with no second mortgage	Census
hunomortgage	Housing units with no mortgage	Census
hdd	Heating degree days base 65	NCDC
cdd	Cooling degree days base 65	NCDC
ipd	Implicit Price deflator 2010	BLS
kwhsite	Electricity per site	Demand Research
gassite	Gas per site	Demand Research
xkwhbill	Total electric expenditures (2010 $)	Demand Research
xkwhcost	Average electric supply cost (2010 $)	Demand Research
xgasbill	Total gas expenditures (2010 $)	Demand Research
xgascost	Average gas supply cost (2010 $)	Demand Research
saveksum	Cumulative kWh savings since 2006	Demand Research
xsavkmeacum	Cumulative kWh ECM costs since 2006	Demand Research
savegsum	Cumulative therm savings since 2006	Demand Research
xsavgmeacum	Cumulative Gas ECM costs since 2006	Demand Research
savkcumratio	Cumulative kWh saving ratio (saveksum/e_usage)	Demand Research
xsavkmeacumratio	Cumulative kWh ECM ratio (xsavkmeacum/xkwhbill)	Demand Research
savgcumratio	Cumulative gas savings ratio (savegsum /g_usage)	Demand Research
xsavgmeacumratio	Cumulative Gas ECM ratio (xsavgmeacum/xgasbill)	Demand Research
college	Number of people with academic college degrees	Demand Research
built2004dum	Census tracts with >30% HU built 2000 to 2004	Demand Research

Variable Name	Variable Label	Source
pge	PG&E	Demand Research
sdge	SDG&E	Demand Research
sce	SCE	Demand Research
d06	Year 2006 Dummy variable	Demand Research
d07	Year 2007 Dummy variable	Demand Research
d08	Year 2008Dummy variable	Demand Research
d09	Year 2009 Dummy variable	Demand Research
d10	Year 2010 Dummy variable	Demand Research

Table C2: Nonresidential Sector Variables (Utility Data Provided by ITRON)

Variable Name	Variable Label	Source
county	County FIPS code	ITRON
tract	Census tract	ITRON
year	Year	Utility
utility	Utility code	Utility
naics	3-digit NAICS code	Utility
elec_sites	Number of electric accounts	Utility
gas_sites	Number of gas accounts	Utility
annual_kWh	Total annual kWh consumption	Utility
annual_therms	Total annual therms consumption	Utility
savings_kWh	ex ante gross kWh annual savings	Utility
savings_kW	ex ante gross kW annual savings	Utility
savings_therms	ex ante gross therm annual savings	Utility
incent_elec	Total kWh incentives	Utility
meaCost_elec	Total electric measure cost	Utility
meaCost_gas	Total gas measure cost	Utility
incent_gas	Total gas incentives	Utility
annual_bill	kWh total annual bill	ITRON
earnings	County industry total earnings (aka personal income)	BLS-REIS
employeecompensation	County industry employee compensation	BLS-REIS
employment	Total county industry employment	BLS-REIS
totalcountyemployment	Total county employment all industries	BLS-REIS
ipd	Implicit price deflator 2010	BLS
kwhsite	Electricity per site	Demand Research
gassite	Gas per site	Demand Research
xkwhbill	Total electric expenditures (2010 $)	Demand Research
xkwhcost	Average electric supply cost (2010 $)	Demand Research
saveksum	Cumulative kWh savings since 2006	Demand Research
xsavkmeacum	Cumulative kWh ECM costs since 2006	Demand Research
xsavkincentcum	Cumulative kWh incentives since 2006	Demand Research
savegsum	Cumulative gas savings since 2006	Demand Research
xsavgmeacum	Cumulative gas ECM costs since 2006	Demand Research
xsavgincentcum	Cumulative gas incentives since 2006	Demand Research
savkcumratio	Cumulative kWh saving ratio (saveksum/e_usage)	Demand Research
xsavkmeacumratio	Cumulative kWh ECM ratio (xsavkmeacum/xkwhbill)	Demand Research
xsavkincentcumratio	Cumulative kWh incent.ratio (xsavkincentcum/xkwhbill)	Demand Research
savgcumratio	Cumulative gas saving ratio (savegsum/g_usage)	Demand Research

Variable Name	Variable Label	Source
xsavgmeacumratio	Cumulative gas ECM ratio (xsavgmeacum/xkwhbill)	Demand Research
xsavgincentcumratio	Cumulative gas incent.ratio (xsavgincentcum/xkwhbill)	Demand Research
xearningsperemp	Per employee earnings (earnings/employment)	Demand Research
xempshare	Employment share (employment/totalcountyemployment)	Demand Research
d06	Year 2006	Demand Research
d07	Year 2007	Demand Research
d08	Year 2008	Demand Research
d09	Year 2009	Demand Research
d10	Year 2010	Demand Research
C1	Office	Demand Research
C2	Retail trade	Demand Research
C3	Educational services	Demand Research
C4	Health care and social assistance	Demand Research
C5	Arts, entertainment, and recreation	Demand Research
C6	Accommodation and food services	Demand Research
C7	All other commercial sector industries	Demand Research
I1	Agriculture, forestry, fishing and hunting	Demand Research
I2	Mining, quarrying, and oil and gas extraction	Demand Research
I3	Utilities	Demand Research
I4	Food, beverage, tobacco	Demand Research
I5	Computer and electronic product manufact.	Demand Research
I6	Nonmetallic mineral product manufact.	Demand Research
I7	Petroleum and coal products manufact.	Demand Research
I8	Plastics and rubber products manufact.	Demand Research
I9	Machinery manufact.	Demand Research
I10	Chemical manufact.	Demand Research
I11	Wood product manufact.	Demand Research
I12	All other industrial sector industries	Demand Research
I13	Construction	Demand Research