Estimating Aggregate Energy Consumption Effects of Energy-Efficiency Programs in California:

Review, Proposed Approach, and Pilot Study

Alan H. Sanstad[*] & Margaret S. Loudermilk[+]

FINAL - August, 2011

Prepared for Itron, Inc. and KEMA, for the California Public Utilities Commission

[*] Berkeley, CA. alansanstad@gmail.com

[+] The University of Chicago, Computation Institute. 5735 South Ellis Avenue, Chicago, IL 60637. Voice: (773) 702-1794. mloudermilk@uchicago.edu

Executive Summary

Over the past decade, the State of California has enlarged its long-standing energy and environmental policy portfolio, including expanded end-use energy efficiency and renewable energy measures and landmark greenhouse gas abatement legislation. Multiple agencies have contributed to policy design and implementation, and numerous specific policies, programs, and regulations have been promulgated in addition to large-scale, over-arching initiatives.

This proliferation of policies and regulations, accompanied by disparate analytical methodologies, has had the unintended consequence of making California's energy-environmental policy and analytical arena increasingly complex and confusing. In the case of energy efficiency regulation, while California pioneered the adoption of rigorous evaluation of publicly-supported efficiency programs, its evaluation infrastructure was not designed for the current era, in which these programs co-exist and interact with a host of other public policies that, directly or indirectly, influence consumers' and firms' decisions regarding energy – particularly electricity consumption.

As part of a larger effort by the California Public Utilities Commission (CPUC), this study focused upon the potential for a new set of aggregate or "top-down" metrics relating to energy efficiency and consumption that would complement the established methodologies for evaluating CPUC's investor-owned utility (IOU) based energy efficiency programs. These metrics would ideally have lower cost than is the case with existing methodologies, as well as enable improved "attribution" of changes in aggregate energy use and/or efficiency between those programs regulated by the CPUC, on the one hand, and other state energy policies and programs under the purview of the California Energy Commission (CEC), the Air Resources Board (ARB), and other authorities, on the other.

To inform the development of such metrics and a quantitative approach for estimating them, a policy and technical background review was undertaken, with the following topics and summary findings:

- *Aggregate energy "intensity" metrics or measures:* These intensity metrics are based on a concept of "efficiency" that cannot directly scaled up to social, demographic, or economic aggregates such as utility service territories, economic sectors, or the state economy. Therefore, such metrics are not suitable for purpose of gauging aggregate impacts of policies or programs to promote energy efficiency or to distinguish such impacts from those of other influences.

- *The current status of California's quantitative energy and environmental policy regulatory infrastructure:* California's energy and environment policy and regulatory infrastructure encompasses multiple simulation models, data sources, loci of authority, statutory constraints, and related factors. Different, often inconsistent and competing, models and methods are applied by different agencies or institutions with overlapping jurisdictions of responsibility. This problem cannot be solved by simply adding a new top-down energy efficiency metric to the mix, regardless of how it is defined and estimated. Simplicity, transparency, and flexibility are therefore high priorities for any new macro consumption metrics.

- *The state-of-development of energy efficiency program evaluation around the U. S :* Efficiency EM&V continues to evolve, but even "traditional" issues such as consistent reporting, treatment of free-ridership, and accuracy are still challenges in practice. The issues addressed in the present study are at the frontier of this field, and there are currently no "role models" to draw upon. Thus, there is an opportunity for California to play a leadership role.

- *The use of econometric and statistical methods in efficiency EM&V:* At the program-specific level, these methods have an established history, particularly in California. They are recognized for their usefulness in *ex post* savings estimation and controlling for free ridership, reflecting the underlying similarity between the measurement and related issues that arise in efficiency evaluation and those that are present in other public policy applications where the methods are used. Because randomized controlled experiments are typically infeasible, alternatives include i) The use of so-called "natural experiments" to identify the effects of policies in the market, for example, the use of control groups of program non-participants in efficiency EM&V; ii) Propensity score matching to correct for simple selection issues – such as energy efficiency free-ridership – when the treated and untreated groups can be matched on a sufficient set of pretreatment characteristics; iii) Instrumental variables, which can be used when a variable exists that is uncorrelated with the outcome of interest but correlated with the treatment (i.e., the policy or program intervention). As in other applications, the core challenge in applying these and other methods is obtaining the empirical data necessary to support econometric or statistical models that can capture the effects of interest, and in designing such models to exploit inevitable limitations, weaknesses, and problems with the data. These data issues are if anything even more difficult in the context of energy consumption and efficiency. Thus, there must be clear understanding and appropriate expectations regarding what this approach can and cannot accomplish, and its usefulness for regulators and other stakeholders.

In summary, our recommendations for macro consumption metrics are:

1. The appropriate top-down metrics are simply absolute energy consumption levels – i.e., in watt-hours and therms (and multiples thereof) – and changes to same due to programs, at any of several scales of aggregation, from utility service territory to statewide sectors (residential, commercial, etc.).

2. These metrics should be estimated econometrically, initially without the direct use of simulation models whether those specific to efficiency program analysis, energy demand forecasting, or statewide economic analysis. The results would be used to estimate energy efficiency program impacts in the aggregate, in the presence of other drivers of consumption, and to assess the feasibility of using this approach for more detailed attribution among energy efficiency policies and programs given the complexity of the regulatory environment and data limitations. Existing program savings estimates should not be directly incorporated in these estimates, which should instead provide a quasi-independent means of gauging aggregate program savings.

3. The problem of consistency, inter-comparison, and coordination of these metrics and their policy applications with other models, methods and measurements is a challenging research problem in its own right, and addressing it should be closely integrated with the development of the metrics.

A proposed pilot study to implement these recommendations would comprise three elements:

1) *An estimation of aggregate, statewide electricity consumption impacts of energy efficiency policies and programs in the residential sector*

A rough estimate of total energy efficiency savings from all sources over the years 2000-2009 will be obtained by performing a linear regression analysis on a panel of geographic aggregates across the United States. The dependent variable will be electricity consumption, and the independent variables will be average electricity price, economic aggregates (average household income, total population, average unemployment, total housing units, etc), weather aggregates (average temperatures or heating and cooling degree days), state, and year. The model will be estimated by standard linear panel methods. Data from California (CEC and CPUC sources), the U. S. Energy Information Administration (USEIA), and the U. S. Census Bureau will be used. The regression estimates will be used to calculate the implied statewide policy effects, with the specification of the model allowing some degree of control for nationwide and intra-state confounding factors, with the estimated effect for California being the combination of state-specific energy policies and any other state-specific effects correlated with energy consumption.

2a) An identification of the programs and/or sets of programs that can be specifically estimated/attributed, the information requirements for doing so, and estimations of their effects (for programs differentiated by time and space within California);

2b) An econometric impact attribution among efficiency programs, codes, standards, and other policy and market influences.

Because comprehensive California micro-level (household) data are neither readily "available" nor currently in a form suitable for this type of analysis, we will instead use information on temporal and spatial differences in efficiency policy and program implementation within California to identify their effects. Some programs are utility or locality specific, while other have been phased in over different periods, and the resulting variation can be used in a similar way to state variation as described in 1), above, to identify program effects. After identifying suitable comparison groups, analysis mirroring that in 1) would be performed to estimate these effects. In the absence of billing data, comparison across CA metropolitan statistical areas from the American Housing Survey may be a viable alternative.

Because of data limitations and the degree of overlap and interaction among programs and categories, it is not possible to directly and definitively estimate or attribute total energy savings among appliance standards, building standards, utility efficiency, and market effects/naturally occurring. However, a coarse overall estimate can be obtained by first estimating the effects of individual programs, aggregating these into categories, and then attempting to control for the overestimate of the total effect in each category that would result from interaction and duplication among the programs and categories. The total amount of this “overlap” might be quantifiable if the effects of all programs were estimated and the total compared to the estimated total California energy savings using the national comparison method previously described. The primary hurdle for implementing this approach is that over 400 separate efficiency programs are currently in place in California. Despite this sheer complexity, as well as the extremely large number of programs relative to the amount and quality of available data, we will attempt to define and apply aggregate program measures.

3) Development and initial testing of an analytical framework for i) Comparing and articulating metric estimates with outputs of one or more existing simulation models used or sponsored by CPUC and/or CEC, and ii) Using metric-based program impact estimates with model-based prospective goal-setting.

In consultation with CPUC and other stakeholders, we will define and implement an analytical framework for structured comparison of a selected set of our econometrically-estimated consumption and savings top-down metric results with those generated by an energy demand model or analytical tool (to be selected) that is in active use in California’s efficiency EM&V and goal setting process. This framework will i) Include technique for taking account of the formal uncertainty quantification associated with the metrics in this comparison; ii) Allow for “harmonization” of the two sets of quantitative estimates without imposing a requirement of equality of numerical values; iii) Integrate metric-based retrospective savings estimates with prospective goals that are model-based.

1: Introduction and Overview

Over the past decade, the State of California has expanded its long-standing leadership role in energy and environmental policy in both scope and scale. End-use energy efficiency has re-emerged as a centerpiece of policy, renewable energy has received renewed emphasis and policy support, and landmark greenhouse gas abatement legislation was enacted in 2006. In each of these areas, multiple agencies have contributed to policy design and implementation, and numerous specific policies, programs, and regulations have been promulgated in addition to large-scale, over-arching initiatives.

This proliferation of policies and regulations, however, accompanied by disparate analytical methodologies, has made California's energy-environmental policy and analytical arena increasingly complex and confusing. Established approaches to quantitative policy analysis have been overtaken, as it were, by the increasing demands placed upon them. This tension is particularly acute in energy efficiency regulation. While California pioneered the adoption of rigorous evaluation of publicly-supported efficiency programs, its evaluation infrastructure was not designed for the current era, in which these programs co-exist and interact with a host of other public policies that, directly or indirectly, influence consumers' and firms' decisions regarding energy – particularly electricity consumption.

This report has been prepared as part of a larger effort by the California Public Utilities Commission (CPUC) to explore, develop, and as appropriate implement, new concepts and methods for i) gauging the effects of its portfolio of investor-owned utility (IOU) based energy efficiency programs, ii) strategic and near-term planning and implementation of new programs, and iii) coordinating and integrating these activities with those of other state energy and environmental agencies and authorities (CPUC 2009). It is specifically focused upon the potential for a new set of aggregate or "top-down" metrics relating to energy efficiency and consumption that would complement the established methodologies for program evaluation currently in place (CPUC 2010b). This investigation is inspired in particular by the work of Horowitz (2011). Key characteristics desired in such metrics are greater simplicity, transparency, ease of implementation, and lower cost than is the case with existing methodologies. It is also hoped that these metrics would enable improved "attribution" of changes in aggregate energy use and/or efficiency between those programs regulated by the CPUC, on the one hand, and other state energy policies and programs under the purview of the California Energy Commission (CEC), the Air Resources Board (ARB), and other authorities, on the other.

In summary, our findings and recommendations are as follows. First, we propose that the appropriate top-down metrics for these purposes are simply absolute energy consumption levels – i.e., in watt-hours and therms (and multiples thereof) – and changes to same due to programs, at any of several scales of aggregation, from utility service territory to statewide sectors (residential, commercial, etc.). Second, we also propose that the appropriate method for estimating these metrics is econometric, initially without the direct use of simulation models whether those specific to efficiency program analysis, energy demand forecasting, or statewide economic analysis. These metrics would be used to estimate energy efficiency program impacts in the aggregate, in the presence of other drivers of consumption, and we would assess the feasibility of using this approach for more detailed attribution among energy efficiency policies and programs given the complexity of the regulatory environment and data limitations. Our approach would *not* incorporate existing program savings estimates; it would instead provide a quasi-independent means of gauging aggregate program savings. Finally, we view the problem of consistency, inter-comparison, and coordination of these metrics and their policy applications with other models, methods and measurements as a research problem in its own right, and propose as part of our pilot study the development and initial testing of a framework for addressing this general coordination problem.

The paper is organized as follows. In the following section, we discuss several key background topics that inform our proposal, including aggregate efficiency metrics, the current state of California's energy policy analysis infrastructure, and econometric energy program evaluation. We then summarize, in Section 3, the fundamental elements of and rationale for our proposed approach. Next, we present a proposed pilot project to implement and test our approach. References are listed at the end of the paper, followed by an appendix with further details on data sources to which we refer in the main text.

2: Background

a. Energy efficiency and energy intensity

Conventional end-use energy efficiency policies and regulations – in California and elsewhere – are based on engineering applications of the first-law thermodynamics applied to specific types of energy-using equipment. The basic principle is that of reducing the amount of energy – i.e., fuel (electricity or natural gas) – required to provide a given quantity of energy *service*, such as space cooling, refrigeration, or lighting.

Characterizing energy efficiency at larger scales has long been approached primarily through commonly-used metrics of energy *intensity* such as energy per dollar of economic output, per capita, per square foot (in buildings), or per unit of output (such as tons of steel or other manufactured products). However, such intensity metrics are neither aggregate versions of engineering efficiency – with, e.g., economic sectors, or statewide building stocks substituting for refrigerators or space cooling systems - nor particularly useful as aggregate measures of the prevalence or penetration of efficient equipment within the larger "entity." Regarding the first point, the core issue is that both the theoretical and the empirical relationships between aggregate energy "inputs" to aggregate economic or other "outputs" are very different from those between fuel inputs to, and energy service outputs from, particular types or units of equipment. Although greater complexity is certainly one aspect of this distinction, it is not the only or even necessarily the most important one. Regarding the second point, the core issue is that there is not, in general, a strong correlation between the penetration of energy-efficient equipment in a sector or other aggregate, on the one hand, and the aggregate level of energy (fuel) consumption in or by that sector, on the other. Other factors – such as (depending on the example) population growth, weather, energy prices and levels of economic activity including household income – can and generally do have greater influence on aggregate energy consumption than the efficiency of installed stocks of energy-using equipment.

Both points are illustrated by the well-known "Rosenfeld curve" comparing per capita electricity use in California, which has been roughly constant since the mid-1970s, with that in the United States overall during the same period, which has continued to rise. This divergence is widely attributed to California's energy-efficiency-promoting policies and regulations, the implementation of which also began in the mid-1970s. There has long been a certain level of controversy regarding this claim, with assertions that, e.g., relatively favorable weather conditions or "structural change" in the state economy accounted for the phenomenon; very recently, the California Energy Commission (CEC) has in fact revised (downward) its estimate of the historical impacts of energy efficiency on the basis of a re-examination of the 1970s and 1980s data on energy savings from utility efficiency programs.[1] For the present discussion, however, consider the CEC's previous estimate and suppose that it accurately reflects the historical impacts of the state's energy-efficiency policies and programs. Then, as Rosenfeld himself has pointed out, these very estimates imply that only around twenty percent of the historical difference between California and U. S. national per capita electricity consumption can be attributed to these policies and regulations (Rosenfeld 2009). We also note that, aside from the magnitude of the "Rosenfeld effect," it is still the case that absolute – not per capita – California electricity consumption has increased substantially over the past four decades. These facts exemplify our previous points: Aggregate per capita electricity consumption is not analogous to equipment-specific engineering efficiency (such as COP or SEER ratings), and absolute aggregate consumption is relatively weakly related to the penetration of end-use efficiency.

Such considerations notwithstanding, it is nevertheless possible to define and estimate aggregate energy intensity metrics. The most thorough and rigorous effort to do so was that of the U. S. Department of Energy in the 1990s (USEIA 1995, Battles and Burns 1998). This study examines and empirically compares energy efficiency/intensity indicators for the U. S. residential, commercial, transportation, and manufacturing sectors, and for the economy as a whole. For example, in the residential and commercial sectors, the indicators include energy per building, per square foot, and per household or employee; in the manufacturing sector, energy

[1] This change has in turn engendered controversy in the context of the California's multi-stakeholder review of historical efficiency program impacts.

per gross output, per value of shipments, and per value of production are among those analyzed. The EIA's findings reflect our previous points on the limitations of aggregate indicators. One of the study's conclusions, for example, is that even within sectors, different indicators can yield results on changes in energy intensity that differ not merely in magnitude but in sign – that is, across sector-specific indicators, there can be uncertainty not only about "how much" energy intensity changes over a period of time, but even about whether it increased or decreased. In addition, the study highlights the inter-sectoral differences among indicators, including that – particularly between the industrial and the other sectors – valid indicators for one sector may be simply indefinable for another.

We conclude that aggregate energy intensity metrics are not surrogates for, analogues to, or estimates of, the narrower physics-based concept of energy efficiency upon which standard public policies are based. This concept cannot be directly scaled up to social, demographic, or economic aggregates such as utility service territories, economic sectors, or the state economy. As a corollary (and for the same reasons), such metrics are not suitable for purpose of gauging aggregate impacts of policies or programs to promote energy efficiency or to distinguish such impacts from those of other influences.

b. California's complicated energy analysis process

The interest in relatively simple metrics stems in part from the complexity not only of the existing program evaluation process but also of the larger analytical infrastructure in California's energy and environment – particularly greenhouse gas (GHG) policy and regulatory infrastructure. There are multiple simulation models, data sources, loci of authority, statutory constraints, and related factors involved. These include the CEC energy demand forecasting modeling system and the ongoing policy process to which it contributes, notably the *Integrated Energy Policy Report*; the ASSET and SESAT models developed by Itron, Inc.; the E-DRAM computable general equilibrium model used by the California Air Resources Board for implementing AB32; the "calculators" developed by E-Three, Inc., and used by CPUC and other stakeholders; the various demand forecasting models and techniques used by the IOUs. While the CEC model and its forecasts have been adopted by the CPUC as "official" for portfolio and program planning, there is otherwise no centralized authority for coordinating these modeling activities, and only limited methodologies for inter-comparison and articulation of their inputs, assumptions, and results. While perhaps more pronounced in California than in other U. S. states, this state-of-affairs is by no means unique in the policy and regulatory world, particularly at the level of national governments: Different, often inconsistent and competing, models and methods applied by different agencies or institutions with overlapping jurisdictions of responsibility.

This problem cannot be solved by simply adding a new top-down energy efficiency metric to the mix, regardless of how it is defined and estimated. Even if such a metric satisfies the criteria of simplicity, relatively low cost, and ease of use – even if it is all-around a "better" analytical tool – it should not be expected, *per se*, to untie the analytical "knot" in California's energy policy and regulation. Indeed, it will be a challenging task to simply determine, quantitatively and with reasonable precision, how metric estimates are related to outputs from existing models, much less how they can be integrated with these outputs.

The example of the Northwest Power and Conservation Council (NPCC) is instructive regarding this point. The NPCC conducts comprehensive supply and demand-side electricity planning for the Pacific Northwest. The NPCC's current overall conceptual policy framework is quite closely aligned with California's in that it emphasizes, among other components, energy efficiency, renewable energy, and CO2 emissions reduction, with the energy efficiency elements being based on the standard principles of cost-effectiveness, efficiency potentials, and so forth (NPCC 2010). While bearing official responsibility and authority for regional power planning, the Council does not directly regulate or oversee individual public or investor-owned utilities. Moreover, the Council maintains and employs a quantitative energy forecasting and modeling system that is at least as complex, in a narrow technical sense, as California's. Yet, as near as we can determine, the NPCC is not subject to the analytical confusion regarding energy forecasting, policy planning, and outcome measurement that is plaguing California. The essential reason, in our judgment, is that there are not *competing* models, methods, or authorities within the domain of the Council's jurisdiction and operations. Thus, its complex models and other analytical methods do not need to be directly coordinated or integrated with others'. This contrast with California is quite fundamentally a matter of different institutional arrangements, not different modeling and analysis methods.

In the same context, we also highlight California's recently-launched multi-stakeholder "demand analysis" project sponsored by the CEC and CPUC, the "DAWG" (Demand Analysis Working Group). This effort arose following the CPUC's adoption of the CEC demand forecast, to address questions and concerns – particularly on the part of the IOUs – regarding the treatment of energy efficiency in these forecasts and in the underlying demand modeling system. *This effort was focused not on developing "new" analysis techniques but rather on understanding those already in use.* That a dedicated, multi-year effort has been required for this purpose is evidence both of the difficulty of the underlying technical issues and of the magnitude of resources and commitments needed to deal with them. We regard the DAWG as an important step in reconciling the complications that have emerged in the quantitative foundations of California energy regulation, and its example has informed our analysis.

c. Efficiency program evaluation across the U. S.

A recent and authoritative review of current practices in efficiency program evaluation, measurement, and verification (EM&V) is presented in Messenger et al. (2010). Rather than repeating or attempting to summarize this review, we highlight several of its findings that are relevant for this project.

First, there is a general although not strict commonality in EM&V approaches and priorities across jurisdictions, including a focus on process, load impact, and cost benefit analyses, an accounting for free-ridership, and the use of deemed savings estimates. Second, there is at the same time considerable variation in key specifics of implementation, including protocols for defining and report program savings, techniques for estimating net (as opposed to gross) savings, and quality control and accuracy. Third, and particularly significant for our study, Messenger et al. find that "…states have made limited progress in addressing the analytic challenges associated with aggregating estimates of gross program savings into load forecasting frameworks." They point out that this problem can in practice "…result in under- or over-counting of savings which may adversely affect resource planning decisions and/or estimates of impacts on greenhouse gas emissions"(*ibid*, p. 29).

The review clearly documents that efficiency program EM&V remains very much a work-in-progress, and that even "traditional" issues such as consistent reporting, treatment of free-ridership, and accuracy are still challenges in practice. In our interpretation, the report's recommendations for EM&V RD&D activities also reveals that the issues being addressed in the present study are at the frontier of this field and that new ideas are needed. This implies that, while exchange of information and coordination with other states on EM&V development may be desirable, there are currently no "role models" for California to draw upon, and that there is an opportunity for California to again play a leadership role.

d. Econometric energy program evaluation

i. Principles and examples

The use of econometric and statistical methods in efficiency EM&V, at the program-specific level, has an established history. Such methods were recognized for their usefulness in *ex post* savings estimation and controlling for free ridership in the California evaluation "Protocols" in the 1990s (CPUC 1998). Examples of their application include Train et al. (1994) and Kandel (1999).[2] This can be seen as reflecting the underlying similarity between the measurement and related issues that arise in efficiency evaluation and those that are present in other public policy applications where the methods are used.

There are several standard econometric approaches to policy evaluation. As in the analysis of medical treatments and pharmaceutical drug efficacy, randomized controlled experiments are ideal. They are, however, exceptionally rare in economic or regulatory policy analysis. One example in the area of energy efficiency is Wolak (2011), which analyzes the effect on electricity demand of a dynamic pricing experiment conducted in the District of Columbia. Wolak estimates the average treatment effect using a linear regression model and finds significant and quantitatively large demand reductions. However, the straightforward casual interpretation

---

[2] Sanstad (2007) provides a background review and analysis of several examples.

presented is not easily generalized since Wolak's empirical example is of a rare, truly random experiment on electricity customers.

A second example of a randomized field experiment in energy efficiency policy is examined in Costa and Kahn (2010). Costa and Kahn analyze data on a field experiment carried out by a large California utility. The utility mailed a Home Energy Report (HER) to a randomly selected treatment group. The HER provided energy usage information relative to that of neighbors along with energy saving tips. Costa and Kahn were able to merge the utility's billing and treatment data with commercially available individual voter registration and marketing data from www.aristole.com . Using this data set they were able to estimate the effect of the HER on household electricity consumption while controlling for certain socio-political factors. Again, it should be emphasized that the causal interpretation is possible due to the random assignment of households into the treatment (i.e., receiving the HER) and control groups.

Because of the difficulty of conducting randomized experiments, a more common alternative is the use of so-called "natural experiments" to identify the effects of policies in the market. This approach has been used extensively in studying policies such as the minimum wage, since adjoining regions may have different policies due to state or local laws. Identification of such as natural experiment can be effective in disentangling the market outcomes of a single policy. In conventional energy efficiency EM&V, the use of control groups of program non-participants is an example of this method. Due to California's leadership role in energy efficiency, many such econometric studies on U.S. energy policy have been conducted using data and examples from this state.

Chong (2011), for example, studies the temperature response of new and old (pre-1970) homes in California. Due to California's extensive energy efficiency building standards, engineering models predict a significant decrease in energy usage in new buildings. However, market responses and other factors could offset these savings. Chong constructs a data set using billing histories for residential households from three IOUs, county assessor data for home characteristic information, U.S. Census data for economic and additional home characteristic information, and temperature data. Using this data set, Chong performs a linear regression analysis of temperature response by home vintage. Without controlling for home characteristics or economic factors, homes built after 1970 are estimated to have a higher temperature response with the newest homes having the largest response. However, newer homes are more likely to have central air conditioning and tend to be larger. In addition, household incomes tend to differ across building vintages. After controlling for square footage, presence of central air conditioning, and income, Chong finds no significant difference between the temperature response of pre- and post-1970 homes. This is still a significant finding since, ceteris paribus, a decrease in temperature response would be expected. Chong is careful to emphasize that no causal interpretation can be made from these results, but he offers many potential explanations, including behavioral and selection effects. Unfortunately, the regression methodology is unable to disentangle these potential contributors in this setting.

Analogously, Jessoe and Rapson (2011) analyze the effect of time-of-use (TOU) pricing by an unspecified public utility for commercial and industrial electricity customers. The structure of this TOU program provided a unique opportunity to potentially econometrically estimate its effect. The utility automatically transferred customers onto TOU pricing if their monthly demand exceeded a specified threshold. After transferring to TOU pricing, they could not return to the regular rate schedule regardless of future demand. All customers had the option to convert to TOU pricing but were also unable to subsequently switch back. Jessoe and Rapson use data on monthly electricity usage, expenditure, rate structure, and whether the customer opted into or was mandated onto TOU pricing. They use a difference-in-difference method to estimate the effect of TOU pricing on customers bills, total usage, and peak load usage. Jessoe and Rapson find that mandatory switchers reduce both their total and peak usage, while voluntary switchers exhibit statistically insignificant increases in both. In spite of having detailed data about participant characteristics and usage, Jessoe and Rapson note two significant issues in using the mandated TOU pricing as an exogenous regulatory treatment. First, if customers altered their demand behavior in expectation of the program's implementation they will underestimate its effect. In contrast, if other policies targeting high-use customers coincided with the implementation of mandatory TOU pricing, they will overestimate the program's effect. Thus, despite a sound methodology in principle, the ability to definitively quantify the effects of TOU pricing in this case are limited.

Reiss and White (2008) estimate the consumption effect of an unanticipated price shock on San Diego customers during the 2000-2001 California energy crisis. By comparing consumption behavior before and after the shock as well as after the subsequent institution of a price cap, Reiss and White use regression analysis to present compelling evidence that average household consumption fell by 13% in response to the price change. However, due to the lack of a true control group, the authors are hesitant to draw a strictly casual interpretation.

A comparable methodology is employed by Jacobsen and Kotchen (2010) in assessing the effect of a change in Florida's energy efficiency building codes. The paper uses residential billing data on both electricity and natural gas along with data characteristics of each residence and compares residences constructed just before and just after the change in 2002 using both simple linear regression and difference-in-difference approaches. Jacobsen and Kotchen find that the increased stringency led to 4% decrease in electricity consumption and a 6% decrease in natural-gas consumption. The key assumption for this result is that unobserved characteristics differ across the two periods that would significantly affect energy consumption. An interesting contribution of Jacobsen and Kotchen's work is a comparison to the results of an engineering simulation model. The predicted increase in energy efficiency from the engineering model was smaller but not statistically significantly different from that obtained in their study. The authors posit that the difference may be due to a combination of spillover effects and a coincident change in appliance standards that are being subsumed in the econometric estimates.

Several studies by Horowitz take advantage of the public availability of aggregate level economic data to estimate the effects of energy efficiency policies relative to a base period. Notably, Horowitz (2007) employs a difference-in-difference approach on aggregate level EIA data to assess the impact of energy efficiency "program commitment" on energy intensity using the Federal Energy Policy Act of 1992 as the beginning of a "treatment period". As noted by Horowitz, this approach requires a highly restrictive assumption that during this period States are otherwise identical to each other in their characteristics and behaviors related to non-program determinants of energy use. For example, this would imply that California, Minnesota, and Mississippi are identical in number of heating and cooling degree-days. Under this assumption though, the differences in the changes in energy intensity across states could be attributed to their program commitment. In an attempt to relax this assumption, Horowitz employs a counterfactual analysis. However, the counterfactual analysis still rests on the assumption that in the absence of differing levels of program commitment the energy intensity trends would be similar. This is a much less restrictive assumption but still of concern if market conditions, consumer preferences, etc differ across states over the sample period. Horowitz (2004) focuses on commercial sector electrical intensity and works to separate the effects of market effects from those of public energy efficiency programs. Although this study differs in many respects from the 2007 study, the pros and cons are analogous.

An interesting intersection of the tradition randomized controlled experiment and natural experiment literatures is found in Allcott (2010a). Allcott examines a pilot program of Chicago's Energy-Smart Pricing Plan in which households that opted into the pilot program were randomly assigned into treatment and control groups. The random assignment controls for household characteristics for households that opted into the program. However, selection into the pilot program among utility participants may still affect results. Thus, as noted by Allcott, this experiment would not be useful in evaluating the effects of a population-wide mandate but does provide consistent estimates for the experimental population (i.e., the opt-in households).

In the absence of a suitable natural experiment, matching or "instrumental" variables may be used to estimate the "treatment effect" of a given policy. Propensity score matching can be used to correct for simple selection issues – such as energy efficiency free-ridership – when the treated and untreated groups can be matched on a sufficient set of pretreatment characteristics. Similarly, instrumental variables are widely used in applied statistical analyses when a variable exists that is uncorrelated with the outcome of interest but correlated with the treatment (i.e., the policy or program intervention). An example in public health research is the use of state cigarette tax rates as an instrument for the health impacts of smoking. However, existing empirical examples of these techniques applied to energy efficiency policy are lacking. This likely indicates a lack of appropriate statistical instruments or characteristics for matching in existing data.

Several other econometric estimation methods exist that are potentially useful in policy evaluation but are currently more a topic of academic research than wide-spread empirical application. These could include estimation of heterogeneous treatment effects, nonparametric estimation methods, and set identification, among others. Allcott (forthcoming) examines the application of heterogeneous treatment effect estimation to energy efficiency policy, specifically the use of Home Energy Report mailings to reduce consumption by twelve U.S. electric utilities. The study finds that households vary substantially in their response to the treatment but estimates an average treatment effect of a 1.1 to 2.8 percent reduction using a difference-in-difference approach.

ii. Practical constraints, data limitations, and their implications

As noted in the introduction, the project of which this study is a part was inspired by the work of Horowitz (2007, 2011) on economic and statistical analysis of energy efficiency programs and policies. Horowitz has made an important contribution not only in his specific numerical results but also in calling attention to the applicability of these methods in this domain as well as their strengths relative to more traditional and established engineering-based techniques, a perspective with which we firmly agree.

As the same time, it is important to correct certain misimpressions that might be conveyed to non-specialists by his discussion of econometric efficiency program evaluation, specifically regarding the ease of applying the techniques and the character of the results that can be expected from doing so (Horowitz 2011[3]). Following are several specific statements that illustrate our concern, and our comments.

> "Econometric models cancel the need for…surveys [to determine free-ridership] because free-ridership is a market effect driven by such variables as energy prices, consumer incomes, and capital stock trends, and properly-specified econometric models control for these effect [sic]" (ibid, p. 45).

Controlling for free-ridership effects is a much more difficult matter than simply "properly specifying" an econometric model. Energy-efficiency free-ridership is an example of the general selection problem that appears in many other contexts, including labor program (i.e., training) analysis and pharmaceutical drug efficacy studies. As in those applications, the core challenge is obtaining the empirical data necessary to support econometric or statistical models that can capture the effects of interest, and in designing such models to exploit inevitable limitations, weaknesses, and problems with the data. As we discuss below, these data issues are if anything even more difficult in the context of energy consumption and efficiency.

[3] This is the paper cited by CPUC (2010b) in its decision calling for this study on 'macro metrics.'

In discussing problems that may arise in correctly attributing demand consumption changes to different factors, Horowitz states (in a particular example) that such a problem would

> "…[point] to two likely modeling problems; (a) either the model is misspecified, meaning that is has omitted variables or suffers from endogeneity bias, or (b) the true demand parameters have changed over the study period for reasons other than energy policy. The first of these problems is easily correctable by revisiting the model. The second problem is also easily correctable" (ibid, p. 48),

and goes on to say that, for the latter correction, "…all that is needed a history of the period under inspection" to determine which government policies may have affected behavior. Regarding the first claim, strictly-speaking, this is true from a theoretical perspective, but in practice the majority of applied econometric work focuses on finding ways to address issues of misspecification, which are not limited to omitted variables and endogeneity. Variables are often omitted due to a lack of suitable data or because they are unobserved, such as customers' preferences for heating and cooling. Similarly, endogeneity is easily corrected when an appropriate statistical instrument is available, but entire literatures have been spawned in debates over such instruments.[4] Regarding the second: The phrase "a history of the period under inspection" similarly obscures the real difficulties that exist both in documenting state-by-state energy policy "histories" with sufficient detail, accuracy, and inter-state uniformity to support any such inferences, and in drawing causal inferences between policy changes and changes in demand parameters, which in itself typically poses econometric data and estimation problems.

In general, Horowitz's claims for the power and ease-of-implementation of the methods he describes are solidly based on an idealized, "textbook" view of the data needed for applying these methods and extremely optimistic assumptions about the ease with which such data can be obtained:

> "…it is not hard to imagine that with reasonable effort, much of these local data can be collected from state agencies such as state energy offices, public utility commissions, and local electric and gas utilities" (ibid, p. 53).

Whatever may be true in other states, we find this statement grossly inconsistent with the realities on the ground, as it were, in California. This is perhaps best described by the comments of the CEC in a recent letter to CPUC Commissioner Grueneich regarding efficiency program evaluation, measurement, and verification:

[4] A canonical example in econometrics is the use of quarter of birth as an instrument for educational attainment in wage regressions. See "Does Compulsory School Attendance Affect Schooling and Earnings?", Joshua D. Angrist and Alan B. Krueger, The Quarterly Journal of Economics, Vol. 106, No. 4 (Nov., 1991), pp. 979-1014.

> "In the course of developing the California Energy Demand 2010-2020 Adopted Forecast (2009), the Energy Commission and CPUC sought to obtain a time-series of IOU program impacts over the last 10 years. This exercise revealed that historical records do not exist in a complete, consistent, comprehensive format. Indeed, it was difficult and time-consuming to reconstruct a record of program accomplishments for just a few years, even at a high level of aggregation. Given the intensity of program reporting EM&V efforts since the inception of the programs, an enormous amount of data exists regarding program accomplishments and the nature of those accomplishments. However, this information is distributed throughout hundreds of EM&V reports and hundreds of regulatory documents – the information has not been compiled in a manner that facilitates comprehensive longitudinal analyses" (CEC 2010b, pp.6-7).

Our review of data resources completely supports these observations. We also note that the situation with California energy consumption and expenditure data – as recorded in customer billing records – is even more problematic. The billing data are not just 'distributed' among utilities, but are also generally proprietary and unavailable (outside the utilities) except in special circumstances.

We therefore find it indeed "hard to imagine" that the data assumed by Horowitz's idealized discussion will be gathered, organized, and made available in California in the foreseeable future. This does not imply that the econometric approach should be avoided in constructing top-down metrics; on the contrary, we will be adopting this approach. It does mean that there must be clear understanding and appropriate expectations regarding what this approach can and cannot accomplish, and its usefulness for regulators: While no magic bullet, it can be an invaluable addition to the policy analysis armamentarium.

3. Proposed metrics and analytical approach

We stated in the Introduction the key elements of our proposal. In this section we re-state and elaborate on these, and discuss in general terms our expected outcomes.

a. The appropriate metrics are watt-hours and therms (and multiples thereof), that is, absolute consumption levels, and changes to same due to programs and/or policies

As we discussed in Section 2, technical energy efficiency is not a meaningful concept at an aggregate level, and various energy intensity metrics have several fundamental limitations for the purpose at hand. In contrast, consumption is meaningful and 'automatically' scalable among different levels of aggregation. It is also comparable across sectors, unlike common energy intensity metrics. This latter property is critical for consistency with "Total Market Gross" as the measure of goals and achievements, as it will allow for consumption changes, including program impacts, to be treated in a uniform and integrated fashion among sectors.

The "complexity" problem discussed in Section 2 also supports the use of absolute consumption metrics, specifically for the purposes of simplicity and transparency. As we noted, the "Tower of Babel" problem, as it were, in California's energy policy and regulatory analysis infrastructure cannot be resolved by simply adding another metric, model, or other tool. However, metrics that are relatively simple and transparent will avoid adding further complexity, and – following the example of the DAWG project – will lend themselves to articulation and comparison with other methods, as described below.

Another important reason for absolute consumption metrics is the overall re-orientation of California energy policy and regulation toward CO2 emissions abatement. At any point in time, aggregate CO2 emissions from the electricity system are a function of consumption levels, not "efficiency." All else being equal, a change in consumption will result in the same-direction change (i.e., increase or decrease) in emissions. This is not the case for either "efficiency" or any energy intensity metric, nor would it be the case for metrics of emissions intensity.

b. An econometric approach is warranted

In Section 2 we reviewed several examples of econometric program evaluation, which illustrated the possibilities of the approach in practice. We propose to extend this type of analysis by estimating our aggregate metrics using "pure" econometric methods.[5] Following is an elaboration of our rationale.

Program evaluation in the open market is always challenging in practice and becomes more so in the case of multiple, overlapping policies or programs. Consequently, there is rarely a single, one-size-fits-all methodological solution. Different methods – including standard "bottom-up" efficiency EM&V - have different benefits and limitations. The key advantage of an econometric approach to aggregate analysis is the potential to control many for socio-economic, market, and multiple policy factors contributing to program uptake, compliance, and outcomes, in a rigorous theoretically-grounded fashion. This capability contrasts with current practice in California. Engineering estimates are clearly better suited and more precise for determining changes in energy output under controlled and reasonably static conditions. However, econometric estimation of actual consumption has been specifically developed to account for, at least partially, the quantitatively important effects of spillovers, free-ridership, and self-selection,

[5] We are to a certain degree over-simplifying the distinction between econometric- and simulation-based models. Many simulation methods make use of classical estimation techniques to some extent, although this is relatively uncommon in energy modeling. Loosely speaking, however, estimation methods compute an exact solution with precisely defined errors (i.e., uncertainties), while simulation models are used to approximate a solution. Approximating a solution may be required when data is sparse or when finding an exact solution is too difficult, and many modern statistical techniques incorporate aspects of both. For the purposes of this report, we will define a "pure econometric" model as one using classical statistical methods for estimation of exact solutions (e.g., regression analysis) without the use of structural models or exogenous engineering information.

as well those of other non-program influences. This is a marked difference not only with engineering methods, but also with simulation models, applied to EM&V.

A pure econometric approach also shares with our proposed metrics the critical characteristic of simplicity and transparency relative to alternative approaches, particularly those involving simulation models. We fully recognize that econometrics and statistics have their own internal complexities, particularly in the dimension of estimation methods. Also, per our remarks at the end of Section 2, limitations in data availability and/or quality – among other factors – introduce intrinsic uncertainty into the results of econometric analysis. However, our opinion is that these considerations are balanced, if not outweighed, by the dramatically greater intelligibility of most econometric models compared to their simulation counterparts. Moreover, a pure econometric approach avoids the use of exogenous assumptions – such as price elasticities of demand – that are a significant source of uncertainty (and dispute) in simulation models. Regarding quantitative uncertainties, these are assuredly present – and typically large in magnitude – but all-too-commonly obscured or ignored in energy simulation modeling. We believe that, with fully-documented models, estimation methods, and data – and with the relevant data made fully available and accessible – our approach will avoid the problems that have arisen (in California and elsewhere) with the perceived “black box” nature of many simulation models.

Finally, we noted at the outset that we would eschew the use of other program savings estimates in our econometric analysis. This criterion will allow us to avoid at least some of the complications and disputes that now surround the estimation of savings in the EM&V process in California. This is possible only with an econometric approach.

c. *Inter-comparison and harmonization of these top-down metrics with other approaches is a challenging research problem in its own right, and needs to be dealt with as such*

As we have discussed, our rationale for an econometric approach to estimating the proposed metrics is strongly influenced by the objective of not exacerbating the current complexity of California’s energy analysis process. Nevertheless, were our approach to be implemented, the metrics – and the formal uncertainty analyses that would accompany them – would need to be quantitatively compared, and ultimately harmonized in some fashion, with the outputs of models such as those mentioned in Section 2b. More generally, the use of the metrics in policy assessment would need to be consistent with the uses of such models. The example of the Demand Analysis Working Group (DAWG) provides strong, if indirect, evidence of both the difficulties in achieving such harmonization and consistency, and the importance of doing so.

It is our view that these problems are comparable in importance and difficulty to those involved in actually estimating the metrics. In fact, they may in a sense be harder, because there is no standard, fully-elaborated methodology to apply. The third component of our pilot project will focus on this issue.

4. Proposed pilot project

In this section we describe how we would implement our approach on a pilot basis. Overall, this pilot would comprise three elements:

1) An estimation of aggregate, statewide electricity consumption impacts of energy efficiency policies and programs;

2a) An identification of the programs and/or sets of programs that can be specifically estimated/attributed, the information requirements for doing so, and estimations of their effects (for programs differentiated by time and space within California);

2b) An econometric impact attribution among efficiency programs, codes, standards, and other policy and market influences.

3) Development and initial testing of an analytical framework for i) Comparing and articulating metric estimates with outputs of one or more existing simulation models used or sponsored by CPUC and/or CEC, and ii) Using metric-based program impact estimates with model-based prospective goal-setting.

These analyses would focus upon the past decade, roughly from 2000 to 2009. Among other reasons, this reflects our understanding of the availability, quality, and accessibility of data, including the program tracking databases created by the CPUC.[6] In addition to program and other California-specific data, federal, state, local and international government agencies along with non-profit and commercial organizations create and make publically available a wealth of socio-economic data potentially relevant to assessing energy efficiency policies. An overview of several of the most important and widely used sources and their applicability to our analysis is contained in the appendix to this report.

The methods discussed in this section are equally applicable in principle to the residential, commercial, and industrial sectors, subject to data availability. Our assessment is that residential sector data are significantly more complete and available. Therefore, in the pilot, the methodology applied to the residential sector, and the results, will provide a test of our approach with the fewest restrictions imposed by data constraints, while the impacts of such constraints will be greater for the commercial sector and greater still for the industrial. Although for convenience the following discussion is phrased in terms of the residential sector, it applies to each sector with the appropriate changes in control variables and/or data sources.

[6] We concur with the CEC staff opinion presented at the May 25, 2011 Demand Analysis Workshop Group workshop, that “There is no reason to believe that re-evaluation of historic efficiency programs will yield more conclusive results due to the lack of adequate ex-post data.”

1: Statewide estimation

A rough estimate of total energy efficiency savings from all sources can be obtained by performing a linear regression analysis on a panel of geographic aggregates across the United States. County level analysis would be preferred. Suitable proxies for economic and weather variables are available at the county level, and the additional observations and associated data variation would increase precision over state level estimates. While California maintains electricity consumption data at a county level through the CEC, similar data for other states are not universally available. Instead, analysis can be performed at the next level of aggregation the utility level, using EIA data. These data come from the US DOE EIA Form EIA-861, the Annual Electric Power Industry Report, and are currently available for 1990-2009. The EIA collects usage and revenues and demand side management program data by utility by sector, and maintains a mapping from utilities' coverage areas to counties. Sales data are in megawatt hours and revenue and expenditure data are in thousand-dollar units. In addition, average utility level prices by sector can be inferred from this data. County level economic and weather data can then be aggregated to the level of the utilities' coverage area and combined to form the panel. A wide variety of county economic data are available from the US Census during the 1990-2010 period (see Appendix). Variables of interest for this analysis would include: population, unemployment, income, housing units, voting data, industry earnings, and business establishments. Weather data are available at extremely fine resolution and can be matched and aggregated in a variety of ways (see Appendix). Indicator variables for each state and year would also be included in the panel. Including indicator variables for each sample year would control for nationwide trends such as the availability of a new electronic device and increasing awareness of environmental issues. State indicator variables would attempt to control for state specific policies and initiatives but would also pick up any state-specific effects not accounted for by the economic and weather controls. The indicator for California would be the primary variable of interest in addition to any demand side management variables included from the EIA data.. If all the determinants of energy consumption are controlled for in the regression equation then the estimated effect of "California" will be the combination of state-specific energy policies and any other state-specific effects correlated with energy consumption. Therefore, if Californians tend to be more environmentally conscious than residents of other states, and this makes more likely to conserve energy and this effect will not be separable from that of specific policies.

Summarizing the above, we say that for each utility, i, in the U.S. and each year, t, from we regress electricity consumption on average electricity price, economic aggregates (average household income, total population, average unemployment, total housing units, etc), weather aggregates (average temperatures or heating and cooling degree days), state, and year. Thus our regression equation is given by

*Electricity_it = a + b_1*price_it + b_2* economic_it + b_3* weather_it + b_4_state_i + b_5*year*

The coefficients are estimated by standard linear panel methods. The coefficient on California will approximate the effect of all state policies on average consumption at the utility level. Using the regression estimates this can be used to calculate the implied statewide policy effects, subject to the caveats above along with those described above.

2: Within-state aggregate program impact estimation (by type), and policy attribution

In general, there are two viable approaches for estimating the effects of specific energy efficiency programs or policies. One is the use of microdata (e.g., household, building, plant, etc. level) in conjunction with indicators of program participation, and demographic, economic, and weather data. California IOU billing data could in principle be matched with a policy/program tracking database, and then matched to census block/track, zip code, county or other levels of aggregated data as previously described. Several examples of this type of analysis were reviewed in Section 2. However, as discussed in that section, these data are neither readily "available" nor currently in a form suitable for this type of analysis. Thus, while this method would be likely to produce the better estimates of the two, it is not feasible for this project. We will therefore take the alternative approach of using information on temporal and spatial differences in policy and program implementation within California to identify their effects.

The feasibility of this approach – i.e., using variation within California across utilities and over time to estimate and attribute program effects - is a potential benefit of the decentralization of California's electricity market. Some programs are utility or locality specific, while other have been phased in over different periods. This variation can be used in a similar way to state variation in step 1 to identify program effects. After identifying suitable comparison groups, analysis mirroring that in step 1 would be performed to estimate these effects. Billing data matched to economic and weather data would be ideal for this analysis to ensure sufficient data variation. In the absence of billing data, comparison across CA metropolitan statistical areas from the American Housing Survey may be a viable alternative.

Currently, energy savings are attributed to categories of appliance standards, building standards, utility efficiency, and market effects/naturally occurring. There is no way to directly and definitively estimate or attribute total savings into these or similar categories. The degree of overlap and interaction between programs and categories precludes this. In addition, the idea that market effects are distinct from policies and programs is misleading at best. Many energy programs indeed affect the relative price of electricity and complementary products such as time-of-use pricing and appliance rebates. However, putting aside the particular category distinctions, it is possible to estimate attribution as follows. The effects of individual programs would first be estimated and then aggregated into categories. Due to the overlap in program effects, this would overestimate the effects of each category. The total amount of overlap might be quantifiable if the effects of all programs were estimated and the total compared to the estimated total California energy savings using the national comparison method previously described.

The primary hurdle for implementing this approach is that over 400 separate efficiency programs are currently in place in California. This sheer complexity, as well as the extremely large number of programs relative to the amount and quality of available data, imply that defining and implementing aggregate program measures will be a major undertaking. This is a key reason for will limiting our focus to the residential sector, i.e., both simple tractability, and – given the previously-noted greater availability of residential data – determining feasibility, or not, for this sector is an obvious benchmark before examining the others.

3: Inter-comparison and harmonization of top-down metrics with existing methods

This component of the pilot will build upon the approach of the DAWG as well as, to a certain extent, that of Stanford University's Energy Modeling Forum. In consultation with CPUC and other stakeholders, we will select an energy demand model or analytical tool – such as Itron's SESAT – that is in active use in California's efficiency EM&V and goal setting process. We will then define and implement an analytical framework for structured comparison of a selected set of our econometrically-estimated consumption and savings top-down metric results with those generated by the model. A critical aspect of this framework will be a technique for taking account of the formal uncertainty quantification associated with the metrics in this comparison, given the essentially fully deterministic character of the existing models; this technique would allow for "harmonization" of the two sets of quantitative estimates without imposing a requirement of equality of numerical values. Another key element of the framework will be the integration of metric-based retrospective savings estimates with prospective goals that are model-based.

We will also explore the possibility of linking the metrics and the model in a simulation-based configuration (see Footnote 7), toward the goal of developing a hybrid approach that more directly integrated the econometric metrics with a deterministic simulation model. Such an approach would hold the possibility of reducing uncertainties in the metrics and ameliorating data limitations, while providing a more rigorous empirical basis for existing models.

References

Cited in the text

Allcott, Hunt. 2009. “Rethinking Real Time Electricity Pricing.” *MIT Center for Energy and Environmental Policy Research Working Paper*, MIT-CEEPR 2009-015.

---. Forthcoming. “Social Norms and Energy Conservation.” *Journal of Public Economics*.

Battles, Stephanie J., and Eugene M. Burns. 1998. “United States Energy Usage and Efficiency: Measuring Changes Over Time.” Paper presented at the 17th Congress of the World Energy Council, Houston, Texas, September 14.

California Energy Commission (CEC). 2010b. “Comments of the California Energy Commission to Follow up Questions on Post-2012 EM&V Issues, Assigned Commissioner’s Ruling, July 2, 2010, CPUC Rulemaking 09-11-014.” Letter dated July 16, 2010, from Energy Commissioners Weisenmiller and Byron to CPUC Commissioner Grueneich.

California Public Utilities Commission (CPUC). 1998. *Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs.* San Francisco, California.

California Public Utilities Commission (CPUC). 2009. *Order Instituting Rulemaking to Examine the Commission’s Post-2008 Energy Efficiency Policies, Programs, Evaluation, Measurement, and Verification, and Related Issues*. Rulemaking 09-11-014, filed November 20.

California Public Utilities Commission (CPUC). 2010b. *Decision on Evaluation, Measurement, and Verification of California Utility Energy Efficiency Programs*. Decision 10-10-033 in Rulemaking 09-11-014, October 28, 2010; Date of Issuance November 2, 2010. San Francisco, California.

Costa, Dora, and Matthew E. Kahn. 2010. “Why has California’s residential electricity consumption been so flat since the 1980s? A Microeconometric Approach.” *National Bureau of Economic Research Working Paper Series,* 15978.

Horowitz, Marvin J. 2001. “Economic indicators of market transformation: Energy efficient lighting and EPA’s Green Lights.” *The Energy Journal* 22, no. 4: 95–122.

-----. 2004. Electricity Intensity in the Commercial Sector: Market and Public Program Effects. *The Energy Journal* 25, no. 2: 115–138.

-----. 2007. “Changes in Electricity Demand in the United States from the 1970s to 2003.” *The Energy Journal* 28 (3): 93-119.

-----. 2011. “Measuring the savings from energy efficiency policies: a step beyond program evaluation.” *Energy Efficiency* 4: 43-56.

Jacobsen, G.D., and M.J. Kotchen. 2010. “Are Building Codes Effective at Saving Energy? Evidence from Residential Billing Data in Florida.” *National Bureau of Economic Research Working Paper Series*, 16194.

Kandel, Adrienne. 1999. “Instrumented Decomposition: A Two-Stage Method for Estimating Net Savings.” *Proceedings of the 1999 International Energy Program* Evaluation *Conference*, Denver, Colorado.

Messenger, Mike, and Ranjit Bharvirkar, Bill Golemboski, Charles A. Goldman, Steven R. Schiller. 2010. “Review of Evaluation, Measurement, and Verification Approaches Used to Estimate the Load Impacts and Effectiveness of Energy Efficiency Programs.” Report LBNL-3277E, Lawrence Berkeley National Laboratory, April.

Northwest Power and Conservation Council (NPCC). 2010. *Sixth Power Plan*. Council Document 2010-09, February.

Reiss, P. C., and M. W. White.. 2008. What changes energy consumption? Prices and public pressures. *The RAND Journal of Economics* 39, no. 3: 636–663.

Rosenfeld, Arthur. 2009. “Green Building: The Performance-Based Approach.” Keynote address at the Symposium on Green Building, the Economy and Public Policy, U. C. Berkeley Program on Housing and Urban Policy and European Centre for Corporate Engagement, Berkeley, California, December 2.

Sanstad, Alan H. 2007. “The Evaluation of Residential Utility Demand-Side Management Programs in California.” Chapter 8 in Richard D. Morgenstern and William A. Pizer, Eds., *Reality Check: The Nature and Performance of Voluntary Environmental Programs in the United States, Europe, and Japan*. Washington, DC: Resources for the Future.

Train, Kenneth, and S. Buller, B. Mast, K. Parikh, E. Paquette. 1994. “Estimate of Net Savings for Rebate Programs: A Three-Option Nested Logit Approach.” *Proceedings of the ACEEE 1994 Summer Study on Energy Efficiency in Buildings*, Asilomar, California.

United States Energy Information Administration (USEIA). 1995. *Measuring Energy Efficiency* in *the United States Economy: A Beginning.* Energy Consumption Series Report DOE/EIA-0555(95)/2, October.

Additional references

Aroonruengsawat, A., and Maximilian Auffhammer. 2009. "Impacts of Climate Change on Residential Electricity Consumption: Evidence from billing data." National Bureau of Economic Research. Cambridge, Mass., USA.

Aroonruengsawat, Anin, and Maximilian Auffhammer, Alan H. Sanstad. 2010. "The Impact of State Level Building Codes on Residential Electricity Consumption." University of California at Berkeley and Lawrence Berkeley National Laboratory. In review.

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

Barker, Terry, Paul Ekins, and Tim Foxon. 2007. "Macroeconomic effects of efficiency policies for energy-intensive industries: The case of the UK Climate Change Agreements, 2000–2010." *Energy Economics* 29, no. 4 (July): 760-778

Borenstein, Severin. 2011. "The Redistributional Impact of Non-linear Electricity Pricing." *Energy Economics* 29, no. 1 (January): 1-17.

California Energy Commission (CEC). 2010a. *Incremental Impacts of the Energy Policy Initiatives Relative to the* 2009 Integrated Energy Policy Report *Adopted Demand Forecast.* Electricity and Natural Gas Committee Report CEC-200-2010-001-CTF, May.

California Public Utilities Commission. (CPUC). 2010a. 2006-2008 Energy Efficiency Evaluation Report. Energy Division, July.

Dilaver, Zafer, and Lester C. Hunt. 2011. "Industrial electricity demand for Turkey: A structural time series analysis." *Energy Economics* 33, no. 3 (May): 426-436.

Ferkingstad, Egil, Anders Løland, and Mathilde Wilhelmsen. 2010. "Causal modeling and inference for electricity markets." *Energy Economics* 33, no. 3 (November 3): 404-412.

Herter, Karen. 2007. "Residential implementation of critical-peak pricing of electricity." *Energy Policy* 35, no. 4 (April): 2121-2130.

Horowitz, Marvin J. 2008. The trouble with energy efficiency indexes: la aritmetic non e opinione. *Energy Efficiency* 1, no. 3 (June 28): 199-210.

Ito, Koichiro. 2010. "Do Consumers Respond to Marginal or Average Price ? Evidence from Nonlinear Electricity Pricing." *Energy Institute at Haas Working Paper Series*, EI @ Haas WP 210.

John, A. 2006. "Modeling Detailed Energy-Efficiency Technologies and Technology Policies within a CGE Framework." *The Energy Journal*, Special Issue #2: 151–170.

Kaufman, Noah, and Karen Palmer. 2010. “Energy-Efficiency Program Evaluations.” Resources for the Future Discussion Paper.

Kaufmann, R.K. 1994. “The effect of expected energy prices on energy demand: implications for energy conservation and carbon taxes.” *Resource and Energy Economics* 16, no. 2: 167–188.

Lin, Boqiang, and Zhujun Jiang. 2011. “Estimates of energy subsidies in China and impact of energy subsidy reform.” *Energy Economics* 33, no. 2 (March): 273-283.

Reiss, P.C., and M.W. White. 2005. “Household electricity demand, revisited.” *Review of Economic Studies* 72, no. 3: 853-883.

Strachan, Neil. 2011. “Business-as-Unusual: Existing policies in energy model baselines.” *Energy Economics* 33, no. 2 (March): 153-160.

Swan, Lukas G., and V. Ismet Ugursal. 2009. “Modeling of end-use energy consumption in the residential sector: A review of modeling techniques.” *Renewable and Sustainable Energy Reviews* 13, no. 8 (October): 1819-1835.

Velthuijsen, J A N Willem. 1993. “Incentives for Investment in Energy Efficiency : An Econometric Evaluation and Policy Implications.” *Environmental and Resource Economics* 3: 153-169.

APPENDIX: OVERVIEW OF DATA SOURCES

a. Economic data

The American Housing Survey (AHS) began in 1973 and collects data on the US housing, including occupant, building, and neighborhood characteristics. National data are collected in odd numbered years, and data for 47 metropolitan areas are collected on a rotating basis. The national sample covers an average 55,000 housing units. Each metropolitan area sample covers 4,100 or more housing units. The survey attempts to both present an accurate representation of the housing market as a cross-section in each survey year and to follow selected housing units over time. Several of California's Metropolitan Statistical Areas (MSAs) are included in the AHS: Anaheim-Santa Ana, Los Angeles-Long Beach, Riverside-San Bernardino-Ontario, Sacramento, San Diego, San Francisco-Oakland (n.b., These become two separate MSAs in the survey around 1998.), and San Jose. The economic, demographic, and building data in the AHS are well suited to econometric analyses of residential energy efficiency policies. Aggregate level data is available on the Internet through the US Census and detailed microdata can be ordered from the Department of Housing and Urban Development at a nominal cost.

The Panel Study of Income Dynamics (PSID) is a longitudinal household study begun in 1968 with a nationally representative sample of 5,000 families. Information on these families and their descendants regarding employment, income, wealth, expenditures, education, and many other topics are continuously collected. In addition to basic socio-economic and demographic controls, the PSID provides many additional variables of interest for EE estimation. Some of these variables include whether the family applied for government energy assistance programs, the type of home heating used, whether the family has air conditioning, and utility expenditures. This extensive data resource is publically available from the Institute for Social Research, Survey Research Center at the University of Michigan.

The Consumer Expenditure Survey (CEX) is maintained by the US Department of Labor's Bureau of Labor Statistics. It provides information on the buying habits of American consumers, including data on their expenditures, income, and household characteristics. The expenditure data includes detailed information on both energy expenditures and expenditures on appliances and other energy consuming devices. Aggregated data is publically available for download from the BLS web site, and the microdata is available for order at a nominal cost. Details of the CEX energy data are discussed further below.

The US Energy Information Administration (EIA) performs three major surveys of energy consumption. These are the Residential Energy Consumption Survey (RECS), Commercial Building Energy Survey (CBES), and Manufacturing Energy Consumption Survey (MECS). The three surveys are closely related in their purpose, scope, and administration. The RECS collects data from a nationally representative sample of housing units on energy use, housing characteristics, energy consuming devices, and some demographic data. This information is combined with data from energy suppliers to estimate energy costs and usage for heating, cooling, appliances and other end uses. The RECS is performed every four years on a small sample of housing units. In spite of the small sample size and limited socio-economic and demographic data, the RECS is notable for the level of detail provided on energy consumption outcomes. Both aggregate level data and microdata are publically available for download. The CBECS is a national survey, conducted every four years, that collects information on the stock of commercial buildings, energy-related building characteristics, energy consumption, and energy expenditures. For the purposes of this survey commercial buildings include structures such as schools, correctional institutions, and buildings used for religious worship. Data of this level of detail for commercial structures is rare, making this an invaluable resource for commercial energy policy evaluation. As with the RECS, both aggregate level data and microdata are publically available for download. However, the in addition to a small sample size and limited set of control variables, major problems have arisen with the CBECS in the last two survey cycles. First, after the 2007 survey the EIA did not release any data or results due to unacceptable data quality. Then the 2011 survey was suspended due to budget cuts at the EIA. It is unclear at this time when the next CBCES may be performed. Finally, the MECS covers the US manufacturing sector with previous surveys conducted in 1985, 1988, 1991, 1994, 1998, 2002, and 2006. Data collection for the 2010 MECS is occurring in 2011, and subsequent MECS are planned to occur every four years. The basic unit of data collection for MECS is a manufacturing establishment. While data on a unique set of variables including participation in energy management activities is collected, it is only publically available at an industry level aggregation (i.e., SIC or NAICS code level). While it is possible that this data could still be used as some form of proxy in econometric estimation, the small sample size and high level of aggregation impose severe limitations.

The data collected by the CEX and RECS are similar but have some important differences, which may be worth a brief digression. The key differences between the surveys, with respect to electricity and energy efficiency data, are that: 1) The CEX collects energy data by surveying respondents, while RECS data is obtained from energy providers; 2) The RECS does not collect data for vacation homes but the CEX does; 3) The CEX and RECS treat renters whose rent payments include energy costs differently. RECS identifies energy expenditures made indirectly through rent but the CEX does not; 4) The CEX apportions energy costs included in homeowner's fees. Instead the RECS works with building management to estimate individual unit costs; and 5) RECS is conducted every 4 years. CEX is an annual survey. The Bureau of Labor Statistics (BLS) has extensively compared the CEX and RECS data for 2001 and 2005. After adjusting the data for these differences to the extent possible, the BLS found that the difference in total expenditures for electricity, natural gas and fuel oils/LPG was less than 10%. In addition the ratio of CEX to RECS expenditures was relatively stable across years for most energy types. CEX estimates were consistently higher with the exception of fuel oil/LPG, however, this may be explained by the small sample size in this category. Thus the CEX and

RECS data appear to be comparable for expenditure measures, and the decision of which data source to use will likely differ by the specific application in question.

In addition to the detailed survey data described previously macroeconomic indicators and aggregate data such as GDP, unemployment, incomes, and production may be desired as control variables in econometric models. These data are available from a variety of sources at national, state, county, zip code, and often even smaller census survey unit levels. The US Census, Bureau of Economic Analysis, and Bureau of Labor Statistics all maintain extensive public data sources.

b. Weather, climate, and related data

One of the most important determinants of the demand for energy is known to be weather. Heating and cooling of residential, commercial, and industrial buildings consumes a large amount of energy. Consequently, to separate the effects of energy efficiency policies from demand effects driven by variations in weather requires data on appropriate climate measures. Surface, air, and water temperatures along with levels of solar radiation and precipitation may all be important controls for an array of policy evaluations. Many of the standard data sources used in geophysical models of earth systems and climate are publically available and easily adapted to use in econometric models. Some of the most definitive sources are briefly described below.

The National Oceanic & Atmospheric Administration (NOAA) maintains the best available precipitation data for the United States in the NOAA Climate Prediction Center .25x.25 Daily US UNIFIED Precipitation data set. The unified data set is derived from three sources that together provide daily precipitation for ¼ degree of latitude by ¼ degree of longitude regions back to 1948. Temperature data to the hourly level ¼ degree of latitude by ¼ degree of longitude regions back to 1979 from the National Centers for Environmental Prediction's Climate Forecast System Reanalysis Project is available from the National Center for Atmospheric Research. Finally, solar radiation data is available from NASA/GEWEX's Surface Radiation Budget project. This data includes daily long and short wave radiation levels on a one-degree of latitude by one degree of longitude grid, beginning in 1983.

Use of these data requires matching the latitude by longitude defined grid squares to the region of interest, typically using GIS (Geographic Information Systems) software. From this point the data may be aggregate to the desired level. Common measures in the energy efficiency literature are the number of heating and cooling degree-days in a given billing month. The level of detail provided by these data allow for the use of many other measures such as the temperature variance or the duration of extreme temperatures.

In addition to the common climate control variables described, evaluation of specific policies may require other related earth system data. Examples might include land cover, soil, or terrain data. For example, in assessing the effectiveness of California's residential building codes, Chong (2011) notes that, "New buildings may differ in their thermal design in that they may have taller ceilings, fewer trees, less passive shading, more structural complexity, or a higher window-to-wall ratio; all of which may increase the electricity needed to cool a building." The inclusion of land cover data might enable a researcher to control for the density of vegetation if this was thought to be an important factor. Integrating data of this type into standard models will typically require extensive processing but is possible and available.