NATIONAL ENERGY TECHNOLOGY LABORATORY

An Investigation into California's Residential Electricity Consumption

July 29, 2009

DOE/NETL-2009/1371

U.S. DEPARTMENT OF ENERGY

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference therein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed therein do not necessarily state or reflect those of the United States Government or any agency thereof.

An Investigation into California's Residential Electricity Consumption

---

DOE/NETL-2009/1371

July 29, 2009

NETL Contact:

Joseph DiPietro
Office of Systems, Analyses and Planning

National Energy Technology Laboratory
www.netl.doe.gov

This page intentionally left blank

Table of Contents

1.0 Introduction 1
1.1 Literature Review 2
2.0 Characteristics of Electricity Consumption 1990-2006 3
2.1 Electricity Consumption per Person (ECP) 3
2.2 Electricity Intensity 8
3.0 Modeling Electricity Consumption in the Residential Sector for 2005 9
3.1 Data 9
3.2 Price Endogeneity 15
3.3 Models 17
3.4 Motivation of the variables 17
3.5 Model I 18
3.6 Matching Estimators 22
3.7 IV Model with California Interaction Variables 23
3.8 Comparison of Results with Other Studies 25
4.0 Conclusion 26

LIST OF TABLES

Table 1 States with lowest ECP vs. States with highest ECP (1990-2006 average) 4
Table 2 States with smallest and largest increases in ECP from 1990-2006 5
Table 3 States with the smallest and largest variances in ECP 7
Table 4 States with lowest electricity intensity (1990-2006) 8
Table 5 RECS Data used in constructing Variables used in Models 10
Table 6 Discrete and Continuous Variables 11
Table 7 Dummy variables 12
Table 8 Distribution of households by Census Division 13
Table 9 Calculation of Variables 14
Table 10 National Model (Model I) Regression Results 20
Table 11 Comparison of IV and OLS Results 21
Table 12 Matching Estimator Model 23
Table 13 CA Interaction Model Regression Results 24

LIST OF GRAPHS

Graph 1 States with the lowest average ECP 1990-2006 4
Graph 2 States with Lowest Electricity Intensities 8
Graph 3 Average cost per kWh vs. kWh consumed for California Homes 16

NETL Contact:
Joseph DiPietro
Office of Systems, Analyses and Planning

Prepared by:

Gavin Pickenpaugh
Office of Systems, Analyses and Planning

Peter Balash
Office of Systems, Analyses and Planning

Acknowledgements

This study would not have been possible without the support of Maria Vargas and Joseph DiPietro, who allowed a significant amount of time to be spent on this research. Additional thanks are due to the reviewers of the paper: Statford Douglas, Lester Lave, James Sweeney, and Dave Harless, who each provided very helpful comments and suggestions. Any errors in this study are those solely of the authors.

1.0 Introduction

This paper investigates apparent discrepancies in electricity usage between the state of California and the rest of the United States. The difference between average electricity consumption per capita (ECP) in California compared to the United States as a whole is a topic of interest in regards to the broad subject of energy efficiency. Since 1970, ECP in California has grown slowly, while ECP in the United States has increased approximately 50 percent. Note that US population increased approximately 46.8[1] percent from 1970 to 2006, while California's population increased 80.8 percent over the same time period. Total electricity consumption in California increased 121.6[2] percent from 1970 to 2006, while it increased 163.6 percent in the United States over the same time period. Thus a simple version of the "California miracle" is that California has not matched the U.S. increase in electricity consumption growth.

The goals of this paper are:

- Examine characteristics of U.S. electricity consumption over time
- Create models to estimate the difference between California and U.S. household electricity consumption after relevant variables have been accounted for
- Develop a model to test whether the effect of factors affecting California electricity consumption is the same as their effect in the rest of the United States.

Section 2 of the paper portrays differences in California electricity consumption through the use of characterizing electricity consumption over time

Breaking down this movement by state would give an indication if there are other states with similar ECP behavior over time as California. Section 2 of this paper examines the movement in ECP from 1990 through 2006 for California as well as other states. During this time period the average ECP for the U.S. (excluding CA) was approximately 12,300 kilowatt hours, while the average ECP for California was 7,000. Thus, California's ECP is approximately 44 percent lower than the U.S. average during this time period.

Section 3 implements Instrumental Variable two stage least square models and a matching estimator model to examine differences in California residential electricity consumption with respect to the rest of the United States.

It would be useful to find out if characteristics affecting U.S. residential household electricity consumption (HEC) are the same as factors affecting California HEC. Section 3 of the paper uses Energy Information Administration (EIA) micro-level data to develop econometric and matching estimator models for household electricity consumption (HEC) in the residential sector. One model includes all states; a matching estimator model attempts to find the nearest matches between U.S. and California households which have similar characteristics other than electricity consumption and location. These models attempt to explain what portion of the difference in HEC between U.S. and California households remains after accounting for other factors. An additional model seeks to explain which factors California households react to differently than households in the rest of the United States.

[1] U.S. Bureau of the Census. (www.census.gov).
[2] Energy Information Administration. (www.eia.doe.gov).

1.1 Literature Review

Other studies have modeled electricity demand. Such a model is complex in that electricity pricing can be non-linear in some areas. It can be argued that customers do not have perfect knowledge concerning electricity pricing.

Sweeney and Sudarhshan (2008)[3] use data from several EIA surveys, Census and other sources to estimate the fraction of the difference between California and the United States electricity consumption per capita that is due to policy independent characteristics such as climate or demographics, and the fraction that may be due to policy measures aimed at saving energy. In the residential sector, variables include income distribution, fuel-mix distribution, household size, urban/rural distribution, and climate in California compared to the rest of the United States. Sweeney and Sudarhshan develop statistics that show what the estimated per household electricity consumption would be for an average US household adjusted to have a particular California characteristic (e.g., household size). For 2001, they find that up to about 23 percent of the overall difference between California and the United States could be due to policy measures, the remainder being explained by various structural factors. The 23 percent is the amount that was not explained as characteristics such as personal taste could not be measured. In the residential sector, Sweeney and Sudarhshan find that up to about 28 percent of the overall difference between California and the United States could be due to policy measures.

Reiss and White (2005)[4] use micro-level data for California from the 1993 and 1997 Residential Energy Consumption Survey (RECS). The study takes into account the two-tier price structure in the California model. Reiss and White develop and estimate a model of household electricity demand that can be used to evaluate alternative tariff designs. This model focuses on the heterogeneity in households' demand elasticities, their relation to appliance holdings and other household characteristics, and how they predict household consumption responses to nonlinear price schedule changes. Reiss and White find that electricity demand can be broken down into an inelastic baseline component, and incremental use (for appliances such as pumps, air conditioners, swimming pools, refrigerators) that is elastic. A mean value for short-run price elasticity is estimated to be -0.4, but as the authors point out, other studies in the residential sector have shown widely varying estimates.

Other studies have attempted to measure the price elasticity of electricity. Riddle performs an overview of the world-wide literature and indicates long-run price elasticity of electricity is typically in the area of -0.7.[5] This can be interpreted as a 1 percent increase in price leads to a 0.7 percent decrease in consumption. Epsey et al. (2004) indicate that the literature widely varies as far as price elasticity of electricity, with price elasticities ranging from 0.076 to -2.5.[6]

---

[3] James Sweeney, Anant Sudarshan. Deconstructing the Rosenfeld Curve. http://piee.stanford.edu/cgi-bin/docs/publications/sweeney/Deconstructing%20the%20Rosenfeld%20Curve.pdf. (Last accessed April 14, 2009).

[4] Peter Reiss, Matthew White. "Household Electricity Demand, Revisited." Review of Economic Studies (2005) 72, 853-883. http://bpp.wharton.upenn.edu/mawhite/papers/Demand.pdf. (Last accessed, April 14, 2009.)

[5] Matthew Riddle. A Review of the Literature on the Price Elasticity of Demand for Electricity. www.e3network.org/ElasticitySurvey2_matt.pdf. (Last accessed April 14, 2009).

[6] Epsey, James; Epsey, Molly. Turning on the Lights: A Meta-Analysis of Residential Electricity Demand Elasticities. Journal of Agricultural and Applied Economics, April 2004.

Mitchell et al. (2009)[7] point to California's "conservation ethic" as playing a factor in their smaller ECP. For example, around 60 percent of California households reported lowering their winter temperature settings when no one would be in the home, while in the U.S. only 45 percent of residents reported taking this measure. Another factor pointing to its lower ECP is that California's manufacturing economy consists of proportionally less energy-intensive industries than the U.S. The study also examined the association between energy efficiency and California's low ECP, but did not find a strong connection.

Kandel, Sheridan, and McAuliffe (2009)[8] use a generalized least squares (GLS) model to estimate California residents use on average 978 kWh less than their weather demographics, and natural gas availability could explain. They estimate that of that remainder, about 40 percent of California's remainder may be explained by policy.

2.0 Characteristics of Electricity Consumption 1990-2006

2.1 Electricity Consumption per Person (ECP)

This section uses EIA state-level data for electricity consumption from 1990-2006.[9] The calculation includes all sectors, while Section 3 of the paper only uses residential data. Data from the Bureau of the Census and from the Center for Disease Control are used to estimate state populations for each year to calculate ECP.[10] [11] An analysis of this data show several interesting characteristics. The state with the lowest ECP over this time period is Rhode Island with 6,899 kWh (California is second). The state with the highest ECP is Wyoming with 25,600 kWh. A breakdown of the five states with the lowest ECP is shown in Table 1.

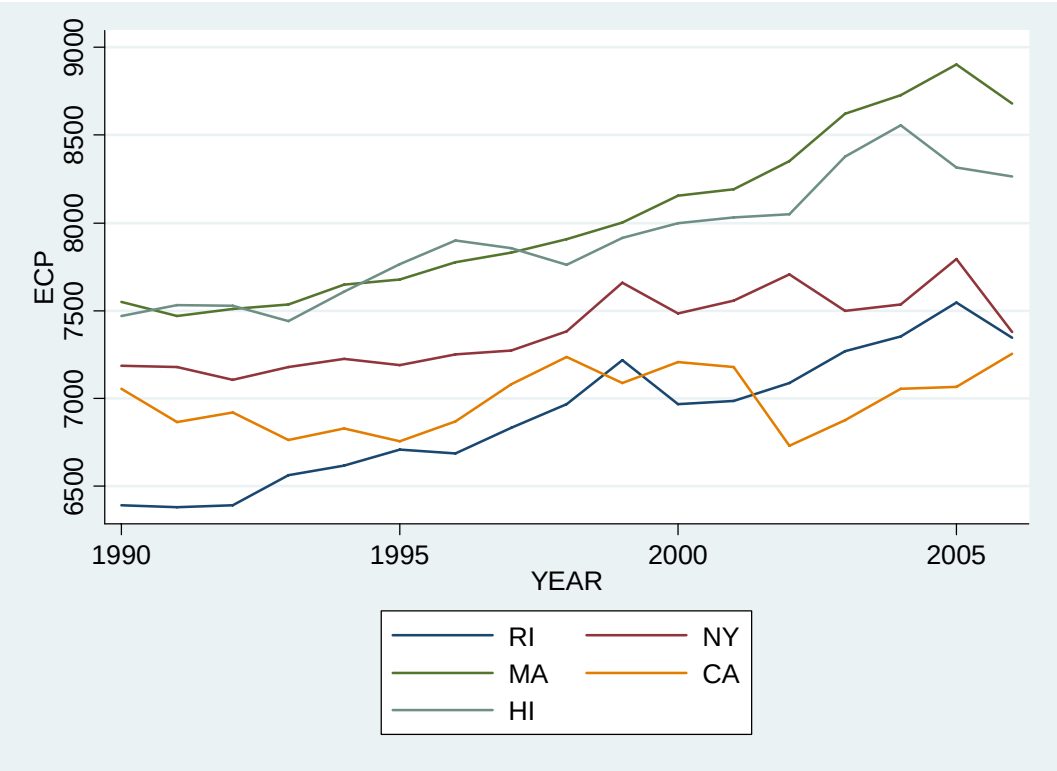
[7] Cynthia Mitchell, et al. "Stabilizing California's Demand." Public Utilities Fortnightly. March 2009.
[8] Kandel, A., Sheridan, M. & McAuliffe, P. "A Comparison of per capita electricity consumption in the United States and California", Paper presented at American Council for an Energy Efficient Economy Annual Conference 2008.
[9] Current and Historical Monthly Retail Sales, Revenues and Average Revenue per Kilowatthour by State and by Sector (Form EIA-826). http://www.eia.doe.gov/cneaf/electricity/page/sales_revenue.xls. (Last accessed January 26, 2009).
[10] 2000-2008 National and State Population Estimates http://www.census.gov/popest/states/NST-ann-est.html. (Last accessed February 17, 2009).
[11] Centers for Disease Control and Prevention. Population Data. http://wonder.cdc.gov/Bridged-Race-v2007.HTML. (Last accessed January 26, 2009).

Table 1 States with lowest ECP vs. States with highest ECP (1990-2006 average)

STATE	ECP (kWh)
States with lowest ECP	
RI	6,899
CA	6,989
NY	7,387
HI	7,904
MA	8,031
States with highest ECP	
ID	16,974
AL	17,413
SC	17,959
KY	19,480
WY	25,600

A time-series plot of each of the states with the lowest ECP (kWh) from 1990 to 2006 is shown in Graph 1.

Graph 1 States with the lowest average ECP 1990-2006



As shown in Graph 1, ECP for Rhode Island, California, New York, Hawaii and Massachusetts trended upward from 1990 to 2006 with varying degrees. California ECP had ups and downs from 1990 to 2002, but increased 7.8 percent from 2002 through 2006 to have an overall upward trend.

Another way to examine ECP movement over time is looking at the percentage change from 1990 to 2006. Using this method California ranks eighth in growth of ECP over the time period. Table 2 shows the eight states with the smallest and largest percentage increases in ECP from 1990-2006. The average percent change in ECP over this time period was 15.9 percent. Notice five of the states with the smallest increases had high ECPs of over 13,000 in 1990, while California, New York, and Maine each had an ECP under 10,000 in 1990. Five states actually had decreases in electricity consumption per capita from 1990 to 2006, with Washington having the largest decrease—28.2 percent. Notice that California and New York had similar percent changes over the 17 year period.

Table 2 States with smallest and largest increases in ECP from 1990-2006

State	1990 ECP	2006 ECP	% Change
States with smallest increases in ECP			
WA	18,578	13,339	-28.20%
OR	15,035	13,023	-13.40%
ID	17,792	15,549	-12.60%
MT	16,409	14,591	-11.10%
ME	9,363	9,343	-0.20%
OH	13,116	13,384	2.00%
NY	7,184	7,377	2.70%
CA	7,054	7,254	2.80%
States with largest increases in ECP			
AL	14,802	19,754	33.5%
MO	10,519	14,049	33.6%
NE	11,305	15,465	36.8%
IO	10,590	14,579	37.7%
WV	12,906	17,865	38.4%
SD	9,092	12,754	40.3%
AR	11,623	16,602	42.8%
ND	11,004	17,641	60.3%

The four states with decreases of over 10 percent warrant further investigation. One thing to notice is that all four states are in the northwest corner of the U.S. The states share many of the same industries, thus a collapse of one industry would affect these states.

The energy crisis in California in 2000–2001 corresponds with marked declines of energy-intensive industries in the northwest. For instance, GDP for the primary metal manufacturing industry fell over 50% in Washington, and GDP for paper manufacturing industry fell over 20% in Oregon, between 2000 and 2003.[12] In 2004, Washington Senator Cantwell sent a letter[13] to President Bush, which included the following statement about industries in the region:

[12] Bureau of Economic Analysis. Regional Economic Accounts. www.bea.gov/regional.

> Energy-intensive industries such as aluminum smelting and pulp and paper production have been driven to the brink of collapse, and the layoffs already number in the tens of thousands. There are innumerable other businesses that are teetering on the brink and simply cannot absorb any more rate increases.

California also showed declines in energy-intensive industries such as paper manufacturing and wood product manufacturing. California's GDP for Wood product manufacturing fell 10.4 percent from 1997 to 2004, while California GDP for Paper manufacturing fell 22.1 percent over the same time period. Between 1990 and 2005, employment in energy-intensive industries in California fell by 20 percent, compared with 16 percent in the rest of the United States.[14]

Electricity prices in the northwest had sharp increases in the industrial sector. In Idaho, from January 2000 to July 2001 average industrial price increased 58 percent; from January 2000 to March 2001 industrial price of electricity increased 167 percent; in Washington from January, 2000 to October 2001 the price increased 94.8 percent; from January 2000 to December 2001 the price increased 59.5 percent in Oregon.

The Rosenfeld curve is the empirical observation that ECP in California has remained relatively flat in California since 1970, while U.S. electricity use per capita has increased around 50 percent over the same time period. In other words, California electricity consumption per capita is stable. A way of measuring stability is by examining variance. One method of doing so looks at each state's variance in ECP from 1990 through 2006. As expected, California's variance is very low, as seen in Table 3. The data were detrended by following a method of removing a linear time trend through the following process:

1. Regress the variable on a time trend and constant
2. Save the residuals

The residuals are linearly detrended and are normalized with a zero mean to allow for comparison between groups. Table 3 shows the states with the highest and lowest variances from 1990 to 2006.

---

[13]Press Release of Senator Cantwell. http://cantwell.senate.gov/news/record.cfm?id=242395. (Last accessed March 13, 2009).
[14] Mitchell (2009).

Table 3 States with the smallest and largest variances in ECP

State	Standard Deviation	Minimum	Maximum
States with smallest standard deviations			
RI	97.4	-126.7	246.3
CO	102.4	-168	156.5
HI	115.1	-151.5	277.4
MA	123.1	-124.0	246.0
CT	123.8	-270.2	229.8
NY	123.9	-283.1	238.5
PA	124.4	-185.5	244.6
CA	159.7	-319.6	248.9
NH	166.8	-272.9	296.9
SD	175.8	-308.7	280.4
States with largest standard deviations			
OH	483.2	-555.5	803.0
OR	490.8	-583.6	963.4
DE	501.1	-1,305.5	1,019.0
SC	517.9	-903	929.2
KY	518.9	-825.5	884.2
LA	593.6	-926.4	836.4
ID	807.0	-908.6	1,390.6
MT	837.2	-1,953.1	1,427.4
WA	970.7	-1,650.3	2,007.3
WY	1,154.8	-1,652.6	2,117.5

2.2 Electricity Intensity

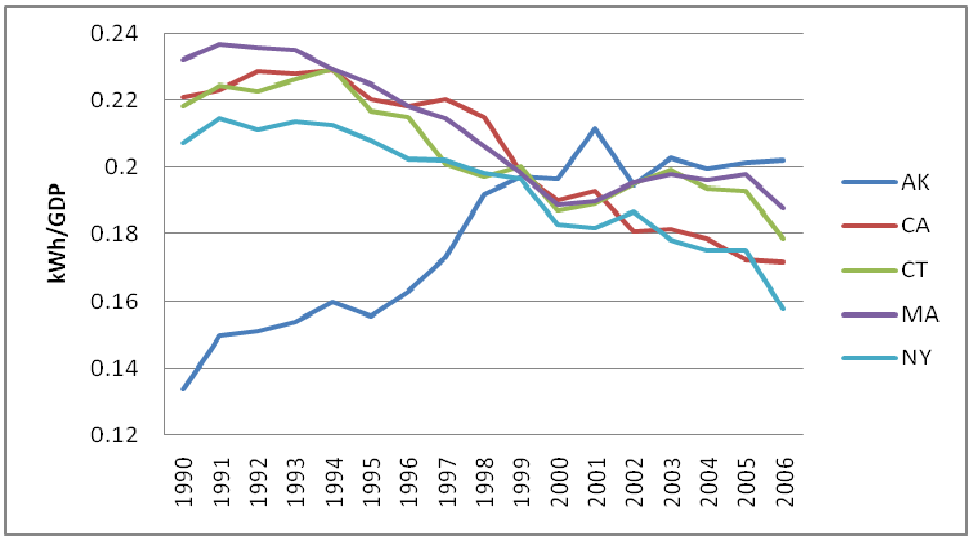
A way to measure efficiency of electricity is taking the ratio of electricity consumption (kWh) to real GDP (Chained 2000$) in different states or the electricity intensity of each state. Table 4 shows that California has the third lowest electricity intensity out of the 50 states from 1990 to 2006.

Table 4 States with lowest electricity intensity (1990-2006)

State	Electricity Intensity
States with lowest electric intensity	
Alaska	0.179
New York	0.194
California	0.204
Connecticut	0.205
Massachusetts	0.211
States with highest electric intensity	
South Carolina	0.689
Alabama	0.697
Idaho	0.714
Wyoming	0.726
Kentucky	0.736

Graph 2 plots the five lowest states from 1990 to 2006.

Graph 2 States with Lowest Electricity Intensities



Four of the five states' electricity intensities have fallen over time, possibly caused by increased efficiency, whereas Alaska has increased from 1990 to 2006, indicating increasing electricity

intensity, perhaps caused by decreased efficiency. When policymakers decide on what constitutes efficiency, it needs to be decided if measures such as GDP should be taken into account or if efficiency should only be measured in energy usage per person.

3.0 Modeling Electricity Consumption in the Residential Sector for 2005

3.1 Data

Data in this section come from the 2005 RECS micro-level data. RECS is a national sample survey of housing units. The survey collects statistical information on the consumption of and expenditures for energy in housing units along with data on energy-related characteristics of the housing units and occupants.

The survey is restricted to housing units that are the primary residence of the occupants; the RECS does not cover vacant housing units, second homes, or vacation units. RECS is conducted by EIA. The RECS was conducted in numerous years between 1978 and 2005. The 2005 RECS used 4,382 housing units. Energy-related characteristics of the housing units and occupants were obtained via on-site personal interviews with the occupants. Energy consumption and expenditures information were obtained from the energy suppliers to the responding households during the Energy Suppliers Survey that followed the household personal interview.

The data obtained from the RECS used in constructing the variables in Section 3.2 is summarized in Table 5. The measures of heating degree days and cooling degree days are not exact.[15] According to EIA, they were altered "slightly" in order to disguise the particular location of a home.

[15] Heating degree days are deviations below the mean daily temperature of 65 degrees Fahrenheit. For example, a temperature of 40 degrees Fahrenheit would correspond to 20 degree days. Cooling degree days are the number of degrees the mean temperature is above 65 degrees for a specific day.

Table 5 RECS Data used in constructing Variables used in Models

Variable	Description	Values
YEARMADE	Gives ranges for when the home was built.	<1940; 1940-49; 50-59; 60-69; 70-79; 80-84; 85-89; 90-94; 95-99; 2000-02; 2003; 2004; 2005
MONEYPY	Gives household income ranges from all sources of income of all members of the households	< $2,500, increments of $2,500 until range $15,000-$19,999, then increments of $5K until $100k-120; $120,000 or more; Unsure; Refused
ELWATER	Indicates whether or not a household uses electricity for heating water	Yes or No
ELFOOD	Indicates whether or not a household uses electricity for cooking	Yes or No
ELCOOL	Indicates whether or not a household uses electricity for air conditioning	Yes or No
ELWARM	Indicates whether or not a household uses electricity for heating the home	Yes or No
LRGSTATE	Indicates whether home is in one of 4 largest states; if so, identifies the state.	none; CA; NY; TX; FL
HD65	Number of heating degree days	Values range from 0 to 11,465
CD65	Number of cooling degree days	Values range from 0 to 5,059
TOTSQFT	Total square footage of the home	Values range from 180 to 9,504
NHSLDMEM	Number that live in the household	Values range from 1 through 15
DOLLAREL	Estimated cost of electricity in dollars	Values range from $56 to $6,670
KWH	Number of Kilowatt hours	Values range from 48 to 72,175
DIVISION	Census Division	New England; Mid Atlantic; East North Central; West North Central; South Atlantic; West South Central; East South Central; Mountain; Pacific

Table 6 shows differences between the US (excluding California) and California in the RECS data used in constructing the models. It also shows other basic statistics such as minimums, maximums, and standard deviations for each category in the U.S. and California. California has much milder climate than the rest of the U.S., larger number of household members, smaller homes in terms of square footage, a higher cost per unit of electricity, and lower HEC.

Table 6 Discrete and Continuous Variables

	kWh	Heating Degree Days	Cooling Degree Days	HH members	Square footage	Cost Elec. ($/kWh)
Average U.S. (excl CA)	11,835	4,625	1,539	2.65	2,344	0.104
Average CA	7,071	1,686	1,044	2.85	1,782	0.123
Percent difference	-40%	-64%	-32%	8%	-24%	18%
Min. U.S.	0	0	0	1	169	0.031
Min. CA	744	591	0	1	167	0.036
Max U.S.	72,175	11,465	5,518	13	11,383	2.23
Max CA	40,340	5,918	4,728	10	8,468	0.302
Standard Deviation U.S.	7,623	2,082	981	1.47	1,658	0.049
Standard Deviation CA	4,860	739	697	1.61	1,182	0.027

Table 7 shows California has an income distribution that includes proportionately more high income and middle income households than the rest of the United States. A smaller proportion of homes in California use electricity for cooling, cooking, and heating water; this has the effect of cutting down on HEC.

Table 7 Dummy variables

Variable	US (excl CA)	CA	Percentage point difference
Income<$20,000	26.8%	19.9%	-6.9
Income $20,000-44,999	31.2%	26.9%	-4.3
Income $45,000-79,999	23.7%	28.0%	6.3
Income >$80,000	18.4%	25.2%	7.0
Veryold (Before 1960)	32.9%	31.4%	-1.5
Old (1960-79)	27.2%	37.4%	10.2
Medage (1980-94)	23.4%	24.6%	3.2
New (1995-2005)	16.5%	6.6%	-9.9
Elfood	64.8%	41.2%	-19.7
Elwater	41.4%	11.3%	-30.1
Elwarm	42.6%	40.4%	-2.2
Elcool	55.4%	43.8%	-11.6

Table 8 shows the distribution of households between Census Divisions

Table 8 Distribution of households by Census Division

Census Division	% of U.S. Total
New England	8.2%
Mid Atlantic	14.3%
East North Central	14.1%
West North Central	7.6%
South Atlantic	13.0%
West South Central	8.5%
East South Central	9.4%
Mountain	8.4%
Pacific	16.4%

The variables were constructed using the RECS data from Table5. A description of how each variable was created is displayed in Table 9.

Table 9 Calculation of Variables

Variable	Formula using RECS data
LHEC	Natural logarithm of KWH
Lcostkwh[16]	Calculated by first dividing DOOLAREL by KWH, then taking the natural logarithm of the result
Lowincome	Equals 1 if MONEYPY is less than $20,000; equal to 0 otherwise
Lowmidincome	Equals 1 if MONEYPY is between $20,000 and $44,999; equal to 0 otherwise
Midincome	Equals 1 if MONEYPY is between $45,000 and $79,999; equal to 0 otherwise
Highincome	Equals 1 if MONEYPY >$80,000; equal to 0 otherwise
Very-old home	Equals 1 if home built before 1950; otherwise 0
Old	Equals 1 if home built between 1950 and 1960; otherwise 0
Medage	Equals 1 if home built between 1970 and 1994
New	Equals 1 if home built between 1995 and 2005
1,000 CDDs	Equals CD65 divided by 1,000
1,000 HDDs	Equals HD65 divided by 1,000
Elwarm	Same as ELWARM
Elcool	Same as ELCOOL
Elc_cdd	ELCOOL * 1,000 CDDs
Elw_hdd	ELWARM * 1,000 HDDs
Lsqft	natural logarithm of TOTSQFT
Lsqftwarm	Lsqft*Elwarm
Lsqftcool	Lsqft*Elwarm
Hhmem	Same as NHSLDMEM
Elwater	Same as ELWATER
Elfood	Same as ELFOOD
CA	Equals 1 if LRGSTATE equals California; otherwise 0
Divrank	Equal to 1 if Home is in Census Division with the lowest price of electricity; 2 if Home is in Census Div. with 2nd lowest electricity price; etc.

[16] This variable measures a household's cost of electricity per kilowatt hour. According to an EIA analyst, the cost of electricity means the cost paid by the consumer, including cost to the utility, taxes, etc.

3.2 Price Endogeneity

The major problem faced in constructing the electricity demand model is that the average price of electricity is endogenous in at least part of the country. This results because in California, there is a non-linear pricing system. This results in a positive relationship between the average cost of electricity and the number of KWH used as shown in Graph 3, which plots this relationship for California households and shows the fit of a locally weighted regression of average electricity cost on number of kilowatt hours. A common solution in the literature to variable endogeneity is to use an instrumental variable. An instrumental variable must be related to the endogenous variable (Lcostkwh), but should not be correlated with the error term. Thus, the instrumental variable should not be part of the electricity demand equation.

The instrument variable examined ranges from 1 through 9, ranking the average electricity price in 2005 in the Census Division in which the household resides. This variable will be called Divrank.

The intuition is that a variable ranking average price for the 9 districts is correlated with the average price per kWh but would not enter into the demand equation. One could argue that prices are affected by that surrounding states and this would affect the price.

Graph 3 Average cost per kWh vs. kWh consumed for California Homes[17]

[17] Lowess smoothing fits a local polynomial regression model to each point and the points close to it. The intent is to provide a picture of the overall shape of the relationship between two variables.

3.3 Models

The models in this paper use 4,382 observations (households) from the 2005 RECS.

3.4 Motivation of the variables

Lcostkwh has an expected negative coefficient; since it is endogenous, it is estimated by using an instrumental variable (Divrank).

Dummy variables were included for four different household income ranges. The motivation for this is that households with more income are expected to use more electricity, holding other factors constant. However, there is the offsetting effect that higher income households may spend more money on more energy-efficient appliances. With low income as the base case, each of the other income dummies is expected to have a positive coefficient, with the coefficient size increasing for each increase in income range.

Dummy variables were included to account for the age of the dwelling. With Veryold as the base case, it is uncertain what the expected coefficient sign will be. Factors such as newer homes having better insulation and other energy-efficient characteristics is countered by the fact that old homes are less likely to have energy-consuming features which may have not been accounted for by the variables Elcool, Elwarm, Elfood, and Elwater. While new homes are likely larger than older homes, the size of the dwelling is already taken into account by the variable Lsqft. Ln(squarefeet) is expected to have a positive coefficient—the larger the home, the more electricity that is needed. Hhmembers is expected to have a positive coefficient—an increase in the number of individuals in a household should increase the household's electricity consumption, holding other factors constant.

To account for climate differences, heating and cooling degree days were combined with Elwarm and Elcool to create interaction terms Elc_cdd and Elw_hdd. The coefficients for Elc_cdd and Elw_hdd are both expected to be positive. A home using electrical cooling (i.e., air conditioning) would be expected to use more electricity than one without. The coefficient for Elwarm and Elcool are included in the model and are expected to have positive coefficients. Lsqftwarm and Lsqftcool both have expected positive coefficients; as square footage increase, a home with electricity used for warming and/or cooling would be expected to use more electricity.

Elwater and Elfood both have expected positive coefficients. Households that use electricity for heating water and households that use electricity for cooking are expected to use more electricity, holding other factors constant.

CA is a dummy variable equal to 1 if the household is in California. The expectation is the coefficient will be negative—a household in California will use less electricity per person than a household that represents the United States as a whole, holding other factors constant.

3.5 Model I

Model I uses two stage least squares[18] instrumental variable (IV) estimation where Lcostkwh is instrumented by Divrank to model the function:

Electricity consumption = *f(Price of electricity, Household income, Age of home, Climate, use of energy-intensive appliances, square footage of home, number of members of the household)*

Assumptions that may be violated include the following:

- **The model is linear in the parameters. This is flexible because the variables can be attributed to functions of the underlying variables of interest, such as natural logarithms.[19]**
- **The error term has an expected value of zero, given any values of the independent variables.**
- **Divrank is not correlated with the error.**

The first assumption could be violated if HEC has a more complicated relationship that OLS cannot model properly. The second assumption could be violated if the functional relationship between the explained and explanatory variables is misspecified. One way this assumption could fail would be if a variable's functional form is misspecified (e.g., using the natural logarithm of a variable, when in fact, the level of the variable should be used). Another factor that can cause this assumption to fail is omitting a variable that belongs in the population model that is correlated with the independent variables. The third assumption could fail if Divrank actually belongs in the electricity demand equation or if it is correlated with the error term.

The results of Model I and their interpretations are displayed in Table 10. Model I has an R-squared of 0.545. This means that 54.5 percent of the total sample variation in the dependent variable is explained by the independent variables. The Wald-test for this model, which tests the joint significance of the independent variables in the model, shows that the variables are jointly significant. The Wald-statistic has a value of 4,937.0, with a p-value of 0.00.[20]

The instrumental variable for Lcostkwh has the expected negative sign (-0.361). In a log-log[21] relationship, the coefficient has an elasticity interpretation. The interpretation in this case is a one percent increase in the cost of electricity per kWh is associated with a 0.36 percent decrease in a household's electricity consumption per person, holding other factors constant.

CA has the expected negative coefficient (-0.254). The interpretation of the coefficient differs in a log-level relationship. In a log-level[22] relationship, the exact interpretation is found by

---

[18] A two stage least squares estimator is an instrumental variable (IV) estimator where the IV for an endogenous explanatory variable is obtained as the fitted value from regressing the endogenous variable on all exogenous variables. See Wooldridge, Jeffrey, M. (2003), Introductory Econometrics: A Modern Approach, 2E.Thomas Southwestern, Ohio.

[19] It is likely that the true conditional expectation (which is never observed) is non-linear, in which case the coefficients cannot be interpreted as marginal effects of changing regressors.

[20] The p-value is the smallest significance level at which the null hypothesis—that the parameters in the model are not jointly significant—can be rejected. For example, a p-value of less than or equal to 0.01 is significant at the 99 percent level.

[21] Log-log means both the dependent and independent variable are in log form.

[22] Log-level means the dependent variable is in log form and the independent variable is in level form.

(100*(e^d-1)% (where d is the coefficient), or in this case (100*($e^{-0.254}$))%. The interpretation of CA is that holding other factors constant, a California household consumes 22.4 percent less electricity than a non-California household.

Hhmem has the expected positive coefficient (0.130). The additional of one household member is associated with a 13.8 percent increase in HEC. Again, the interpretations of dummy variables and level variables23 use the formula (100*(ed-1)% (where d is the coefficient) because they have a log-level relationship with the dependent variable.

The income groups each have the expected positive coefficient. Since low-income is the base case, each income dummy is being compared to low-income. The interpretations are a low-mid-income household consumes 9.2 percent more electricity than a low-income household; a middle-income household consumes 23.1 percent more electricity per person than a low-income household; a high-income household consumes 33.8 percent more electricity per person than a low-income household.

Old and Medage are each individually insignificant; however, an F-test of joint significance for Old, Medage, and New results in a p-value of 0.02, so the age of the home is jointly significant. One peculiarity in the results is that the coefficient for New is negative while the coefficients for Old and Medage are positive. This could be partly attributed to newer homes being more likely to have more energy-efficient appliances than older homes.

For Elwarm and Elcool to be evaluated at their means, the means of Lsqft, Lcdds, and Lhdds had to be used in constructing Lhddwarm, Lcddcool, Lsqftcool, and L sqfootwarm.24 Elwarm and Elcool both have the expected positive coefficients. Lhddwarm and Lcddcool also have the expected positive coefficients, and both are significant, with a p-value less than 0.01. Lsqftwarm and Lsqfootcool also have the expected positive coefficients. An increase in heating (cooling) degree days for a home that uses electricity for heating (cooling) is associated with an increase in HEC.

The results of Model I suggest a California household uses 22.4 percent less electricity per person than a non-CA household. In reality (without holding other factors constant), a California household consumes 40.3 percent less electricity than a U.S. household. The unexplained portion could potentially be explained by factors such as different behavior to changes in variables among U.S. and California households (e.g., differing price elasticities). The model in Section 3.7 attempts to quantify these differences by including California interaction variables.

Table 11 compares the IV results to OLS if an instrumental variable was not used. Note the large difference in the coefficients for Lcostkwh in IV (-0.361) and OLS (-0.656). Thus, a smaller electricity demand electricity is estimated when using the IV method. The decrease in R-squared is not large from using IV (0.557 versus 0.545 for OLS).[25] One result consistent in both OLS and IV models is that the coefficient for California is negative and is statistically significant. Thus, even if the IV used is not valid, the combination of results provides evidence that after accounting for other factors, California still has a significantly lower HEC than the rest of the United States.

---

[23] Wooldridge (2003) notes that the expected log used in the formula is upwardly biased, but is a consistent estimator.

[24] For example, Lhddwarm equals Elwarm x (Lhdd minus the average of log of heating degree days).

[25] By definition, the R-squared for OLS is always lower; sometimes the difference can be large in magnitude. See Wooldridge (2003) for further discussion.

Table 10 National Model (Model I) Regression Results

Dependent variable: LHEC		
Independent Variable	**Coefficient**	**Interpretation**
Lcostkwh(IV'd by Div)	-0.361** (0.061)	1% increase in cost of electricity is associated with a 0.36% decrease in HEC
Lowmidinc ($20,000-44,999/yr)	0.088** (0.021)	Lowmidinc household uses 9.2% more electricity than lowincome household
Midinc ($45,000-79,999/yr)	0.208** (0.022)	Midinc household uses 23.1% more electricity than lowincome household
Highinc (>$80,000/yr)	0.291** (0.026)	Highinc household uses 33.8% more electricity than lowincome household
Old (1960-79)	0.007 (0.019)	Home built between 1960-79 uses 0.7% more electricity than one built before 1960
Medage (1980-1994)	0.014 (0.021)	Home built between 1980-94 uses 1.4% more electricity than one built before 1960
New (1995-2005)	-0.053* (0.023)	Home built between 1995-2005 uses 5.1% less electricity than one built between before 1960
Elwarm	0.169 (0.018)	Home with average # HDDS and square footage using electricity for warming uses 18.4% more electricity than home not using electricity for warming
lhddwarm	0.044 (0.015)	If home uses electricity for heating, 1% increase in heating degree days is associated with a 0.04% increase in HEC
Lsqftwarm	0.070 (0.019)	1% increase in a home's square footage is associated with a 0.07% increase in HEC for home using electricity for warming.
Elcool	0.313 (0.019)	Home with the average # of CDDs and square footage using electricity for cooling would use 36.7 % more electricity than a home that did not use electricity for cooling
Lcddcool	0.220 (0.021)	For home using electricity for cooling, 1% increase in cooling degree days is associated with a 0.22% increase in HEC.
Lsqftcool	0.212 (0.018)	1% increase in a home's square footage is associated with 0.21% increase in HEC for home using electricity for cooling.
Hhmem	0.129** (0.005)	An increase of one household member is associated with an 13.8 percent increase in HEC.
Elwater	0.341 (0.019)	Home using electricity for water heating uses 40.7 percent more electricity than one that does not.
Elfood	0.075 (0.018)	Home using electricity for cooking uses 7.8 percent more electricity than one that does not.
CA	-0.254 (0.027)	Calif. home uses 22.4% less electricity than one in the rest of the U.S.
Constant	7.379** (0.136)	
# Observations: 4,346 R-squared = 0.545		****Significant at 99% level *Significant at 95% level**

Table 11 Comparison of IV and OLS Results

Independent Variable	IV Estimation	OLS
Lcostkwh	-0.361**	-0.656**
	(0.0614)	(0.0505)
Lowmidinc($20K-45K)	0.0877**	0.0948**
	(0.0208)	(0.0203)
Midinc($45K-80K)	0.208**	0.226**
	(0.0222)	(0.0217)
Highinc(>$80,000/yr)	0.291**	0.319**
	(0.0260)	(0.0251)
Old(1960-79)	0.00711	0.000784
	(0.0194)	(0.0192)
Medage(1980-1994)	0.0141	0.0105
	(0.0207)	(0.0204)
New(1995-2005)	-0.0526*	-0.0565*
	(0.0228)	(0.0225)
Lhddwarm	0.0439**	0.0261
	(0.0148)	(0.0147)
Elwarm	0.169**	0.142**
	(0.0177)	(0.0170)
Elcool	0.313**	0.273**
	(0.0186)	(0.0177)
Lcddcool	0.220**	0.230**
	(0.0215)	(0.0210)
Lsqftwarm	0.0703**	0.0630**
	(0.0186)	(0.0183)
Lsqftcool	0.220**	0.215**
	(0.0176)	(0.0177)
Hhmem	0.129**	0.127**
	(0.00507)	(0.00509)
Elwater	0.342**	0.316**
	(0.0187)	(0.0182)
Elfood	0.0752**	0.0608**
	(0.0175)	(0.0171)
CA	-0.254**	-0.219**
	(0.0271)	(0.0279)
Constant	7.379**	6.746**
	(0.136)	(0.112)
# Observations	4346	4346
R-squared	0.545	0.557
**Significant at 99% level	*Significant at 95% level	

3.6 Matching Estimators

Model I attempts to address, other things equal, how much California HEC consumption differs from that of other states. Stated differently, is there a California "treatment effect"? A way to answer the question and a way to have a comparison to Model I's results is to use a matching estimator. The logic is that for variables one wants to hold constant (income categories, age of home categories, etc.), one matches a household in the "treated" group (California households) to one or more members of the "untreated" group (other U.S. households). An algorithm is used to find nearest matches and then estimates the "treatment effect" by examining the difference in the dependent variable between the matched households. This method has the advantage of avoiding issues surrounding specifying the "correct" functional form.

Matching estimators[26] attempt to impute the missing potential outcome by using average outcomes for individuals with "similar" values for the covariates. The matching estimator model in this section uses the nnmatch command in Stata. This estimator estimates the average treatment effect on the dependent variable (kWh) by comparing the outcomes between treated and control (CA and non-CA households), using nearest neighbor matching. The program pairs observations to the closest matches in the opposite treatment group to provide an estimate of the treatment outcome. Exact matching is often used for binary variables. In this case four matches were used and exact matching was used for the following variables: Elwarm, Elcool, Elwater, Elfood, Lowmidinc, Midinc, and Highinc. This gives more weight to those variables to ensure that they are exactly matched.[27]

Table 12 summarizes the results of the matching estimator model. The results of the matching-estimator models (30.9 percent) estimate a larger California effect than Model I (22.4 percent), but this is expected since price of electricity is not accounted for in the matching estimator models.

[26] For a detailed treatment on types of estimators used in this model reference Guido Imbens et al., 2004. "Implementing matching estimators for average treatment effects in Stata," Stata Journal, vol. 4(3), pp. 290-311. For a background of psmatch2 estimators, reference Austin Nichols, 2005. "Causal inference with observational data." State Journal, vol. 7(4), pp. 507-41.

[27] Old Medage and New were not given an exact command because from the earlier regression results they are less important determinants of electricity consumption.

Table 12 Matching Estimator Model

Type of Matching estimator (Stata command)	Coefficient	Interpretation	% Change from U.S. (excl. CA) Average
Matching estimator (NNMATCH)	-3,167.8*	For households in the treatment group, the average effect of residing in California is an annual decrease in HEC by 3,167.8 kWh	-30.90%
***Significant at 99% level**			

3.7 IV Model with California Interaction Variables

The use of interaction variables allows for the estimation of the differences in California's parameters and the U.S. The model is similar to Model I except in addition to each variable in Model I (with the exception of CA), identical variables are created that are multiplied by CA. This results in the following general equation;

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \beta_{0c} CA_i + \beta_{1c} x_{1ci} + \beta_{2c} x_{2ci} + \ldots \beta_{kc} x_{kci} + \varepsilon_i$$

The interpretation of the coefficients is as follows:
β_j is the coefficient for the U.S. outside of California. β_{jc} is the difference between the coefficient of x_j for California and the coefficient of x_j for the rest of the country. The sum of the two coefficients is the full California coefficient.

Table 13 summarizes the results of the model. The far right column sums β_{jc} and β_j to show the full California coefficient. The sum of β_j and β_{jc} variables show that California households have a higher coefficient (in absolute value) for every variable except Elc_cdd, Hhmem and the intercept. A test of joint significance of the variables with interaction terms rejects the null hypothesis that the differences between the California and U.S. variables is jointly 0 (p-value = 0.00).

Variables which have significantly different coefficients for California compared to the rest of the U.S. include Lsqftwarm, Lsqftcool, Elwarm, Lhddwarm, and Lcddcool. These estimate associate changes in square footage, HDDs, and CDDs with different changes to electricity consumption for California and the rest of the U.S. The California interaction income groups and home age dummy groups are not jointly significant.[28][29]

[28] An F-test for joint significance of the income dummies for California results in a p-value of 0.241, while the same test for home age dummies results in a p-value of 0.289.
[29] Unlike in model I, Elwarm and Elcool were not evaluated for a home with 0 square feet and 0 HDDs or CDDs; this resulted in the large negative coefficient—intuitively, a model would not be able to estimate a home with 0 CDDs/HDDs that is 0 square feet, since it cannot exist.

Table 13 CA Interaction Model Regression Results

Dependent variable: LHEC			
Variable	**B_{jc}**	**B_j**	**B_j+B_{jc}**
LcostkWh (IV'd)		-0.376** (0.0626)	
Lowmidinc ($20-45K)	0.069 (0.075)	0.086** (0.022)	0.1539
Midinc ($45-80K)	0.138 (0.071)	0.194** (0.023)	0.332
Highinc (>$80K)	0.0672 (0.080)	0.281** (0.027)	0.3482
Old (1960-79)	0.0647 (0.062)	0.004 (0.0205)	0.06857
Medage(1980-94)	-0.058 (0.064)	0.0177 (0.0218)	-0.0396
New (1995-2005)	0.035 (0.092)	-0.0498* (0.024)	-0.0144
Lsqftwarm	0.115* (0.059)	0.0561** (0.020)	0.1711
Lsqftcool	0.193** (0.067)	0.214** (0.018)	0.407
Elwarm	-2.85** (0.689)	-0.477* (0.188)	-3.328
Elcool	-0.746 (0.716)	-2.857** (0.240)	-3.603
Lhddwarm	0.260** (0.084)	0.029 (0.016)	0.2887
Lcddcool	-0.081 (0.067)	0.219** (0.023)	0.1381
Hhmem	-0.020 (0.0155)	0.132** (0.005)	0.1124
Elwater	-0.003 (0.081)	0.337** (0.019)	0.33356
Elfood	0.098 (0.053)	0.059** (0.019)	0.1571
Intercept	-0.35** (0.0819)	7.357** (0.137)	7.009
#Observations: 4,346	**99% level of significance		
R-squared: 0.553	*95% level of significance		

3.8 Comparison of Results with Other Studies

As mentioned in Section 1.1, Espey et al. (2004) indicate that the literature widely varies as far as price elasticity of electricity, with price elasticities ranging from 0.076 to -2.5. Reiss and White (2005) find a short-run price elasticity of -0.4. The results of Model I fall within these ranges with an estimated U.S. electricity price elasticity of -0.361.

Model I estimates that after accounting for other factors, California HEC uses 22.4 percent less than HEC than the rest of the United States. The actual difference in the sample used in the model is approximately 40.3 percent; therefore, somewhere around half of the difference may be explained after controlling for other variables. In the residential sector, Sweeney and Sudarhshan (2008) find that up to about 28 percent of the overall difference between California and the United States could be due to policy measures. They find that the other 72 percent is explained by adjusting the average U.S. household to have the same characteristics as a California household. Kandel, Sheridan, and McAuliffe (2009) estimate California residents use on average 978 kWh less than their weather demographics, and natural gas availability could explain. They then estimate that of that remainder, about 40 percent of California's the remainder may be explained by policy. The results of this study suggest that these figures may be underestimating the effect of policy and other unexplained factors on electricity consumption.

4.0 Conclusion

This paper has examined factors concerning the differences in California electricity consumption relative to the rest of the United States. Some key points from the paper are:

- California is well below the national average in electricity intensity; however, it is neither the least electricity intensive state nor has it had the lowest growth rate in ECP since 1990.
    - While many states have had large increases in ECP over time, California ECP increased less than 3.0 percent from 1990 through 2006.
- Average electricity price cannot be used in an OLS model because for certain states (e.g., California), it increases with the level of consumption and is thus an endogenous variable in the electricity demand model.
- The results of the models indicate several key points:
    - The results of Model I estimate that when holding constant factors such as cost of electricity, members per household, income, climate, housing age, California households use about 22.4 percent less electricity than those in the U.S. as a whole. This suggests that somewhere around half of the difference may be attributable to differences is other factors, such as policy, individual tastes, and other unknown factors.
    - Factors that are significant in predicting electricity consumption are income, average cost of electricity, size of the home, climate, and whether the home uses electricity for heating, cooling, cooking, and water.
    - The results of matching estimator model, which does not take the cost of electricity into account, estimates a household in California uses somewhere around 30.9 percent less electricity than a household with similar characteristics in the rest of the United States.[30]
    - The estimates from the model with California interaction terms suggests that variables that may have significantly different effects between California and the rest of the U.S. are square footage, number of household members and whether a home uses electricity for cooking.

A component that could provide a more complete picture would be to quantify how much energy efficiency programs influence electricity consumption in the residential sector. Horowitz (2007) found that states with utility efficiency programs saw reductions in electricity use and intensity in the commercial and industrial sectors compared to states with little or no efficiency programs. However, the study did not find decreases in electricity consumption in the residential sector.

A future area of study could follow that by Reiss and White (2005) by collecting individual rate schedules and matching them to specific homes and then using this data to calculate the price elasticities for each home. The difficulty in implementing this approach would be from collecting rate schedules from all utilities in every state as well as the fact that EIA made small changes to cooling degree days and heating degree days in order to prevent individual households from being identified.

[30] The matching estimator model does not take average electricity cost into account because of the endogeneity problem.