One of These Homes is Not Like the Other: Residential Consumption Variability

Phillip Kelsven

Itron

October 2013

ABSTRACT

Households consume energy in many different ways for many different purposes. While many households use energy in similar and predictable ways, the majority of homes use energy in unpredictable ways that lead to great variability in the total annual energy consumption. The behavior patterns in which American households use energy causes wide variations in total residential energy consumption per home, even when normalizing for size and similar determinants. Data suggest that even in homes that are similarly sized, the variability in energy consumption can vary by factors of eight or higher. What accounts for such wide variations in energy consumption in fairly homogeneous households?

This study relies on research conducted with available data from electric and gas utilities in two different states. The goal was to calculate the variance in energy consumption among groups of homogeneous households. Data sources include utility consumption data, house characteristic data from county tax assessors, and U.S. Census-based demographic data. To perform the analysis, first homes were grouped into segments of similar sizes, vintages, and demographics to create homogenous groups. The variance in total energy and heating energy was then analyzed to learn of the magnitude of the difference among housing segments. Second, the research sought to identify possible demographic and housing stock indicators using regression modeling that cause greater variability in total observed energy consumption. Third, a methodology for further research at the household level was developed to explain the individual variations in consumption.

This research has implications for marketing successful energy efficiency programs. Households who appear remarkably similar often have vastly different reasons for installing energy efficiency measures in their homes. There is no one size fits all approach. Thankfully, enhanced customer engagement strategies are receiving renewed focus among energy efficiency programs nationwide. This paper will describe how other sectors of the economy including retail, insurance, health, and politics are using big data to develop segmented marketing approaches. Further, it will develop guidelines for using the methodology in this paper to conduct marketing research to provide valuable insight on residential customers' apparent energy efficiency behaviors and related decision making. This information, in turn, will help to make energy efficiency program outreach more effective.

Readers of this paper will learn just how much energy consumption can vary among homogeneous housing stocks and demographics, the factors that appear to influence large variations in energy consumption, and inferences on how energy consumption affects program participation. Attendees will benefit from a line of research that is not widely available and a new methodology to study the topic of energy consumption variability. In addition, they will be taught how crafting a different message that resonates with each segment may be a more

effective marketing strategy. Finally, attendees will be presented with a methodology to do this type of analysis for their own customer base.

Introduction

Organizations working to promote energy efficiency must gain a better understanding of variability in energy consumption, variability in rates of energy efficiency program participation, and the factors that affect these differences in order to more-efficiently achieve energy savings in a greater number of homes. This paper explores the variance of energy consumption among homes of similar sizes and socioeconomic characteristics. This paper also explores how the variance relates to participation in energy efficiency programs, and how socioeconomic factors may affect variability in energy consumption and energy efficiency program participation. Data from homes in two electric and two gas utilities in two different states are assigned into fairly homogenous segments whose distribution of energy consumption is analyzed. These segments are then compared to each other for differences in consumption distributions and in energy efficiency program participation. This paper attempts to explain differences between segments with socioeconomic indicators from the US Census.

Methods - Forming Homogenous Segments

Data from 140,000 gas-heated homes in a western state and 70,000 homes in an eastern state are compiled from utility consumption data, house characteristic data from tax assessors, and US Census data. To make use of the data for this study the exact location and name of the utilities must remain anonymous but shall hereafter be referred to as the western and eastern samples. All of the homes in the samples are single family, gas-heated homes whose square footage and year built have been obtained from county tax assessor data. As a result, observations for some variables are at the home level, while observations of others, such as proportion of college graduates, are at the census block group level. For each area, homes are grouped into relatively homogenous segments on the basis of house size, vintage, and neighborhood demographics. The western sample contains homes in the same weather zone in three neighboring Counties. The eastern sample contains homes throughout the same state but in the same weather zone.

Home size was considered the most important factor in creating homogeneous segments. Homes were first segmented into nine blocks based solely on house size at 300 square foot intervals, however the largest home size categories are larger than 300 square foot intervals. Homes that were larger than 4,500ft^2 or smaller than 1,000ft^2 and 700ft^2 respectively were discarded, as this extreme characteristic prevented them from belonging reasonably to any homogenous group. Each of the nine blocks were then further segmented by a series of statistical cluster analyses based on income, vintage, educational demographics, ethnic demographics, homeowner proportions, neighborhood house type proportions (e.g.-single family, mobile home), the proportion of homes in the neighborhood considered to reside in an urban setting, and age demographics. Special attention was taken to ensure that homes were clustered into segments with houses of similar expected energy efficiencies based on vintages. Homes that did not fit well into any homogenous group, due to extreme values of one or more characteristic, were excluded from the final segmentation. The final segmentation produced 362 homogenous segments in the western sample and 307 segments in the eastern sample.

Normalized Annual Consumption (NAC) estimates of gas and electricity consumption were prepared from monthly meter reads provided by the household's utility providers. The NAC procedure estimates the gas and electricity consumption in a typical weather year for each home using regression techniques against heating and cooling degree days calculated from the nearest weather station. 2009 data was used to estimate household energy consumption. Total energy consumption (in thousand BTUs) was calculated from gas and electric utility consumption data. Households that lacked electricity consumption data were discarded for this part of the analysis. This reduced sample sizes considerably for some of the eastern segments. Two segments had fewer than 30 observations containing electricity data and were considered inadequate for analysis involving electricity and total energy consumption. Because all homes in the sample are gas-heated, all observations included gas data. Individual households with gas, electric, or total consumption that fell more than five standard deviations from their segment's mean were also excluded from the respective analyses to prevent misrepresentations of segment consumption distributions.

Energy Consumption Variance Within Segments

Energy consumption was found to vary significantly among the homogeneous housing segments created. A consumption "spread" statistic was developed to analyze the difference in consumption in each segment from the 5^{th} to the 95^{th} percentile of consumption in gas, electric, and total energy consumption. Removing observations outside of the 5^{th} and 95^{th} percentile, or essentially cutting off the tails of the distribution curve further eliminates outliers from influencing the analysis. The spreads for high and low variance segments are reported in tables 1 and 2. The average spread of total BTU energy consumption in the western sample is 3.00; in the eastern sample it is 2.83. This means that similar sized homes with similar household demographics have total energy consumption that varies by a magnitude of about 3. If outliers were included such as the minimum and maximum consumption homes per segment the spread ranges from around 6 in the low variance segments to as high as 16 in the high variance segments. More detailed data on the spread of consumption is found in the appendix.

Table 1: Summary of Consumption Spread by Energy Type in Western Sample

Energy Type	Mean Spread from 5^{th} to 95^{th} Percentile
Gas 5 Lowest Variance Segments	2.90
Gas 5 Highest Variance Segments	4.26
Electricity 5 Lowest Variance Segments	4.26
Electricity 5 Highest Variance Segments	5.67
Total Energy 5 Lowest Variance Segments	2.59
Total Energy 5 Highest Variance Segments	3.88

Table 2: Summary of Consumption Spread by Energy Type in Eastern Sample

Energy Type	Mean Spread from 5th to 95th Percentile
Gas 5 Lowest Variance Segments	2.70
Gas 5 Highest Variance Segments	8.50
Electricity 5 Lowest Variance Segments	5.02
Electricity 5 Highest Variance Segments	7.65
Total Energy 5 Lowest Variance Segments	2.37
Total Energy 5 Highest Variance Segments	5.82

The spread in energy consumption within segments differs depending on the energy type and location. Electricity has more variation than Gas in both samples. This makes sense since the stock of appliances and home electronics can differ significantly between similar sized homes, whereas every home has similar needs for space heating. There is more variation in gas consumption in the eastern sample than the western sample. The western sample has a milder climate compared to the eastern sample so it makes sense that gas consumption shows more variation in the east. This evidence suggests that energy using household behavior is more diverse with electricity end uses including appliances, lighting, and electronics than it is with gas heating end uses, suggesting that electricity is where a majority of behavioral savings may exist.

It is important to note the shape of the consumption distributions. Skewness is a measure of the asymmetry of a distribution and indicates whether deviations from the mean tend to be positive or negative. The skewness indicates whether a majority of the observations lay to the right or left of the mean. For the western households, the distributions are all positively skewed, or skewed to the right, with the majority of observations falling to the left of the mean. Examples of different kinds of distributions found in this analysis are shown in figures 1-3. The distributions are also all leptokurtic, with narrower peaks and fatter tails compared to the normal or bell shaped distribution. The degrees of skewness and kurtosis vary by segment and by energy type. The eastern data shows similar trends, although the average skewness and kurtosis in these segments are significantly lower and the range in skewness is greater, with some metrics even depicting slight negative skewness. This suggests that the eastern sample tends to have a more normal looking distribution whereas the western sample tends to have a majority of homes in the lower end of the consumption distribution. Skewness and Kurtosis tests for normality indicate that a majority of the western sample energy consumption distributions as well as the eastern sample gas consumption distributions depart significantly from normality (α=.05), although no conclusions regarding normality can be confidently made about some of the segments.

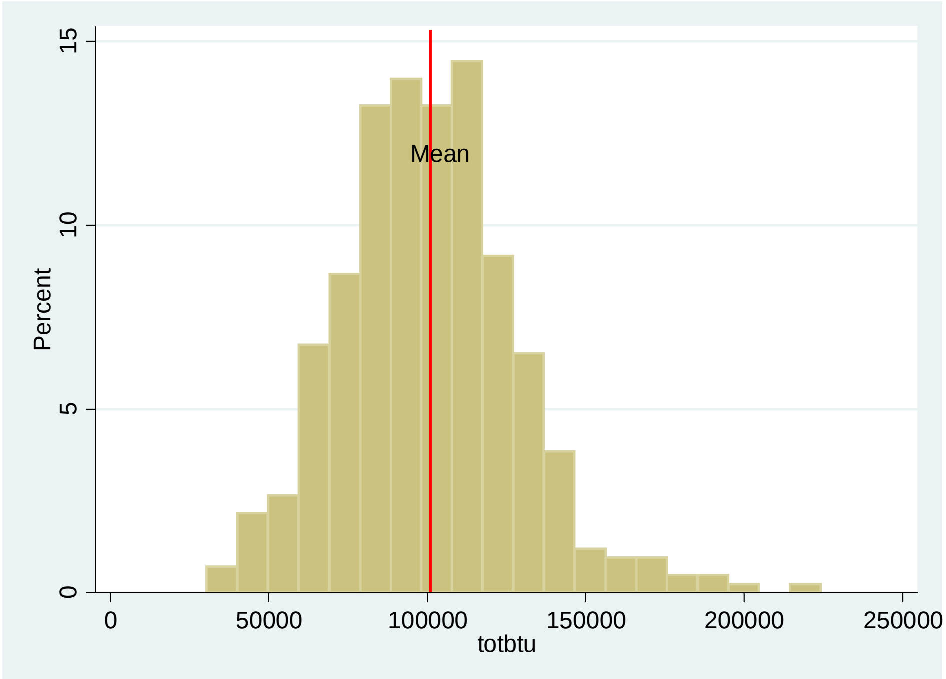

Figure 1: Slightly Positive Skewness

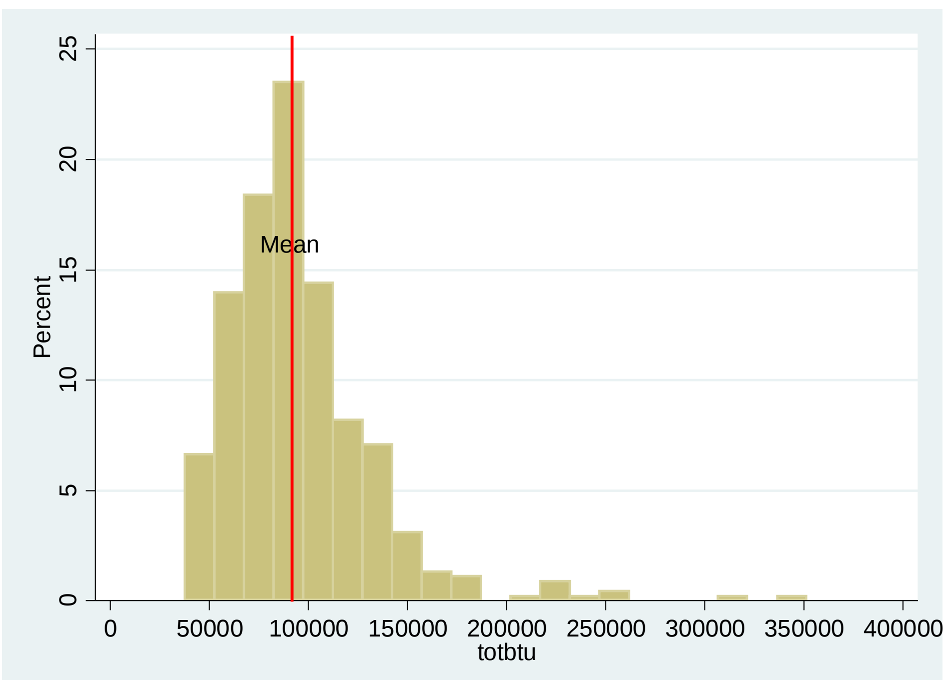

Figure 2: Large Positive Skewness

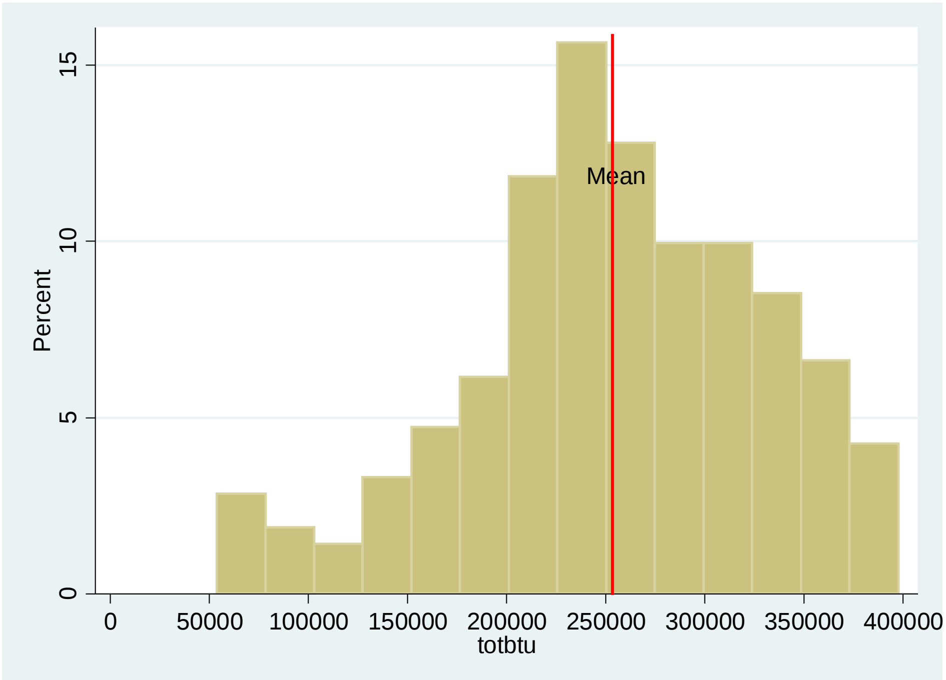

Figure 3: Negative Skewness

Energy Consumption Variance Across Segments

This study finds that the variances of energy consumption between each segment vary significantly for all three types of energy at each geographical location. This is a significant finding that indicates homogeneous segments do not appear to have similar distributions to each other. Each housing segment displays unique distribution characteristics. In terms of total energy consumption, the Bartlett test for equal variances, the Levene test for equal variances, and the Brown-Forsythe test statistic demonstrate statistically significant variation in consumption standard deviation across segments as shown in table 3.

Table 3: Significance of Variation in Consumption Standard Deviations

State	Energy Type	Test Statistic Type	Test Statistic Value	P-Value
Western Sample	Gas Only	Bartlett	14,000	<.001
		Levene	18.84	<.001
		Brown-Forsythe	16.55	<.001
	Electricity Only	Bartlett	11,000	<.001
		Levene	6.25	<.001
		Brown-Forsythe	5.21	<.001
	Total BTU (Gas + Electricity)	Bartlett	20,000	<.001
		Levene	16.97	<.001
		Brown-Forsythe	14.64	<.001
Eastern Sample	Gas Only	Bartlett	16,000	<.001
		Levene	39.54	<.001
		Brown-Forsythe	38.04	<.001
	Electricity Only	Bartlett	3,700	<.001
		Levene	10.72	<.001
		Brown-Forsythe	9.47	<.001
	Total BTU (Gas + Electricity)	Bartlett	7,600	<.001
		Levene	22.04	<.001
		Brown-Forsythe	21.23	<.001

Evidence suggests that the variance is greater between segments for gas than electricity consumption. As shown in Table 1, tests for equality of variances show that gas consumption varies more significantly between segments than electricity consumption due to larger test statistics. Combining this with the finding that electric consumption has a greater spread suggests that electric consumption has a somewhat consistent variance between segments, but gas consumption for heating is where a greater difference is evident. This may be due to significant differences in building envelope as well as behavioral elements.

The eastern sample consumption appears more diverse than the western sample energy consumption. Table 1 depicts more significant differences in variances of gas, electricity, and total energy consumption among the eastern segments than among the western segments, presuming normality is not assumed. It is unclear whether this difference is due to climactic factors, structural factors, or behavioral factors. The colder winter climate in the eastern sample no doubt plays a role in this. It is also noteworthy that all tests for equal variances using all energy types at both geographical locations are significant at the 0.001 level.

Homes in homogenous segments with the highest variations in energy consumption tend to consume a lot more energy, on average, than segments with the lowest consumption variations. They tend to include a larger proportion of homes that use extremely large amounts of energy.

Homes that display the highest variation expressed in standard deviations also tend to be large homes. As tables 2-13 in the appendix indicate, four of the five segments with greatest total energy variance in the western sample are segments larger than 2,800 square feet. All five of the segments with the greatest total energy variation in the eastern sample are greater than 3,000 square feet. It follows that homes that rank among the lowest variation in total energy consumption are smaller homes. Four of the five western segments with low variance are less than 1,600 square feet, and all of the low variance homes in the eastern sample are less than 1,300 square feet.

It must be noted that there exists variation in segment sample size that could potentially affect the standard deviations in the various energy types per segment. Larger samples are often associated with smaller variances as measured by standard deviation. Segment sample sizes range from a low of 38 to a high of 1,977. This issue was fully researched and variance in segment sample size does not bias the analysis in a significant way. There are many examples of segments with large sample sizes that have significantly larger standard deviations in energy consumption.

Socioeconomic Factors and How They Relate to Energy Consumption Variance

This study finds that significant variation in standard deviations of energy consumption can be explained by a number of socioeconomic factors. To test which factors help explain differences in variation US Census data at the block group level was appended to the housing segments. This data is not at the household level, but nonetheless is an indicator of neighborhood demographics and housing stock makeup. Econometric models are used to show how socioeconomic factors may explain differences in consumption variance between housing segments.

To determine the best models for each of gas, electricity, and total energy consumption in each of the two areas, many ordinary least squares (OLS) models were created using all possible combinations of available independent variables that did not violate the OLS BLUE (Best Linear Unbiased Estimator) assumptions. Robust standard errors were used in all of these models to account for heteroskedasticity, and all models controlled for sample size. The best models, as defined by the models with all independent variables significant, high goodness of fit (r-squared) and the lowest Akaike Information Criteria (AICs), were chosen from this exhaustive set of models. For these purposes, any two variables with a correlation coefficient, r, of 0.5 or greater were considered to be highly correlated.

Six regression models were developed to explain the energy consumption variance for three energy types in the western and eastern samples. The models explaining standard deviation in each of gas, electricity, and total energy consumption from the Oregon data are strong, with between 54% and 78% (r-squared of .54 and .78) of the variation in standard deviations explained by the chosen independent variables. The models from the eastern data were even stronger, with the chosen independent variables explaining 78-88% of the consumption variation. The full set of regression model results can be found in the appendix.

We will focus on the models that explain variance in total energy consumption in BTUs. Certain home characteristics help to explain consumption variance. Home size is positively related to increases in consumption variance, although this relationship is much stronger in the eastern sample. The proportion of mobile homes is weakly negatively correlated with variance in both models. Older homes appear to show increasing variance in energy consumption in both models which is unsurprising as more recent building standards for new homes are more rigorous ensuring more homogeneity in consumption. Additionally, some older homes have been significantly weatherized and retrofitted with energy efficiency measures causing significantly lower heating and cooling consumption than homes that are less weatherized. Neighborhoods in the western sample with a greater proportion of single family housing units are associated with lower energy variance. This analysis is limited to consumption in single family homes, so this finding suggests that neighborhoods that are more exclusively single family are more homogenous in energy consumption.

Household demographics show significant correlations with energy consumption variance primarily only in the western sample, where home square footage and year built are the only variables with significant power to explain consumption variance in the eastern sample. Among the western households, neighborhoods with greater proportions of high-income households are associated with increasing energy consumption variance suggesting that some higher income households use a lot of energy, while some do not. Income is weakly correlated with increasing variance in the eastern sample but is highly correlated with home size which has more explanatory power. Housing segments with older homeowners in the western sample tend to be positively correlated with increasing variance. This relationship is very weak to non-existent in the eastern sample.

Table 4: Regression Model for Total Energy Consumption Variance in western sample

R-Squared - .77

Indicator	Relationship	T-Statistic	P-Value
Home SQFT	Positive	6.11	<.001
% Mobile Homes	Negative	-2.20	.030
Mean Year Built	Negative	-4.26	<.001
% Single Family	Negative	-8.86	<.001
% Income > $100k	Positive	20.12	<.001
Median Age	Positive	2.23	.026

Table 5: Variables Left Out of Western Model

Indicator	Reason
% Home Ownership	Correlated with % Single Family, % Income > $100k, Median Age
% College Educated	Correlated with % Income > $100k, Median Age
Household Size	Correlated with Median Age

Table 6: Regression Model for Total Energy Consumption Variance in eastern sample

R-Squared - .88

Indicator	Relationship	T-Statistic	P-Value
Home SQFT	Positive	35.30	<.001
% Mobile Homes	Negative	-2.69	.007
Mean Year Built	Negative	-13.68	<.001
% Urban	Positive	2.30	.022

Table 7: Variables Left Out of Eastern Model

Indicator	Related To
% Single Family	Not Significant
% Income > $100k	Correlated with Home SQFT
% College Educated	Not Significant
Household Size	Correlated with Home SQFT
% Home Ownership	Not Significant

While skewness also varied substantially across segments, this variation cannot be explained well by the socioeconomic demographics used in this study to explain variation in standard deviation. Several models using the same variables that are used in the variance models were experimented with, but all suffered from a poor model fit.

Energy Efficiency Program Participation Analysis

Similar residential energy efficiency programs were conducted in the utility territories of the homes in this analysis. Some residents chose to participate in their local program, while others abstained or missed the opportunity. We seek to understand how energy efficiency program participation relates to variance in energy consumption as well as consumption levels and socioeconomic indicators. For each homogenous group, a participation rate was calculated. This metric is the proportion of homes that participated in the program in some manner divided into all households who are eligible for the program per census block group.

Participation rate appears to be significantly positively correlated with variance in total energy consumption in Massachusetts and variance in gas consumption in Oregon. Controlling for sample size only, variance of gas consumption in the western sample and variance of gas, electric, and total energy consumption in the eastern sample are significantly and positively correlated with participation rate according to simple regression analysis of all housing segments. This indicates that participant homes tend to come from segments with larger consumption variation. This suggests that neighborhoods with high program participation have households with diverse consumption and behavioral patterns, at least more diverse than neighborhoods with lower participation rates.

Participants versus Non-participants

Program participants in the highest participating housing segments do appear to display unique consumption characteristics from non-participants in those segments. We analyzed the median consumption and consumption standard deviation of the segments with the ten highest number of program participating households and find that participating households in the western sample tend to have higher electric, gas, and total consumption than non-participant households. Seven of the top ten participating segments in the western sample have participants with higher total energy consumption than non-participants. These differences, however, are small and not statistically significant. The lack of statistical significance is likely due to large standard errors resulting from the relatively small sample of participants compared to non-participants per segment.

Contrary to findings in the western households, program participants in the eastern sample tend to consume less total energy than their non-participant counterparts. Eight of the top ten participating segments in the eastern sample have participants with lower median electric and total energy consumption than non-participants, and five of ten have lower gas consumption. This suggests that the western households that use greater than average amounts of energy are more likely to participate in energy efficiency incentive programs, whereas the high users in the eastern sample are less likely to participate in energy efficiency incentive programs. The above average energy users in the western sample tend to want to save energy and the eastern households that are already using less energy than average for their housing segment want to save more.

Differences between participants and non-participants in energy consumption variance depend on type of energy and geography. The variance of consumption among participants is greater for gas consumption in the western sample and there are no clear differences in the variance of electricity and total energy between participants and non-participants. Program participants in the eastern sample tend to have lower variances in all three types of energy. Differences in energy consumption variance between participants and non-participants can only be established with statistical significance for electricity consumption in the eastern sample. As depicted in Figures 4 and 5, the distributions of electricity consumption among the top ten participant segments for participants and non-participants differ slightly. The difference is seen in the right tails of the distributions where the non-participants have a fatter tail, meaning more non-participant households use large amounts of electricity. Participants in the eastern sample have smaller variances in electricity consumption and use slightly less electricity on average than non-participants, and few households at the high end of electricity consumption appear to participate in energy efficiency programs.

The overall trend suggests that housing segments with higher participation rates tend to have greater consumption variance for all types of energy in the eastern sample, and for only for gas consumption in the western sample. Focusing on the participants versus non-participants tells a somewhat different story however. The western participant's consumption variance appears to be larger than non-participants only for heating energy use. The western program participants have similar electric consumption distribution to the rest of the population suggesting that significant inefficiencies in space heating may be the driver of participation in these housing segments, and that behaviors in electricity use are similar between participants and non-participants. Program participants in the eastern sample have lower consumption variation among all energy types than non-participants. This goes against the finding that higher participation is positively related to consumption variance. What accounts for the discrepancy in the eastern sample? The answer is in the non-participant energy variance. The non-participants in the high participation segments in the eastern sample have significantly larger variance in consumption and the participants are actually more homogenous in consumption. The non-participants in these high participation segments in the east are using a significantly more energy than their neighbors who install energy efficiency measures.

Table 19 – Differences in Electricity Consumption Variance Among Participants and Non-Participants in the Top Ten Participant Segments in Eastern Sample

Segment	Participants	Non-Participants	Standard Deviation of Consumption (Participants, kWh)	Standard Deviation of Consumption (Non-Participants, kWh)	Bartlett Test Stat (chi^2)	Levene Test Stat (F)	Brown-Forsythe Test Stat (F, W50)
93	62	502	2,782	3,789	8.804 (p=.003)	4.730 (p=.030)	4.297 (p=.039)
17	61	428	3,040	3,499	Insig	Insig	Insig
103	57	431	4,033	3,820	Insig	Insig	Insig
72	53	537	2,710	3,479	5.515 (p=.023)	3.914 (p=.048)	Insig
175	52	250	3,891	4,117	Insig	Insig	Insig
140	50	335	3,387	3,827	Insig	Insig	Insig
43	49	360	2,909	3,224	Insig	4.500 (p=.035)	4.338 (p=.038)
44	46	369	2,713	3,530	4.810 (p=.028)	Insig	Insig
149	46	351	2,952	3,931	5.647 (p=.017)	4.636 (p=.032)	4.369 (p=.037)
170	45	133	3,650	3,835	Insig	Insig	Insig
All 10	521	3,894	3,347	3,775	12.448 (p<.001)	6.50 (p=.011)	5.60 (p=.018)

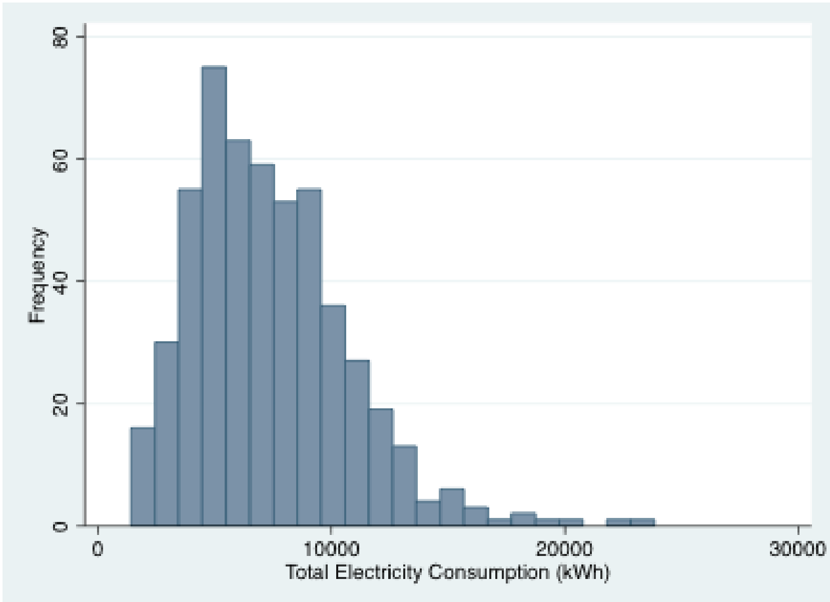

Figure 4 – Distribution of Electricity Consumption Among Participants in the Top Ten Participant Segments in Eastern Sample

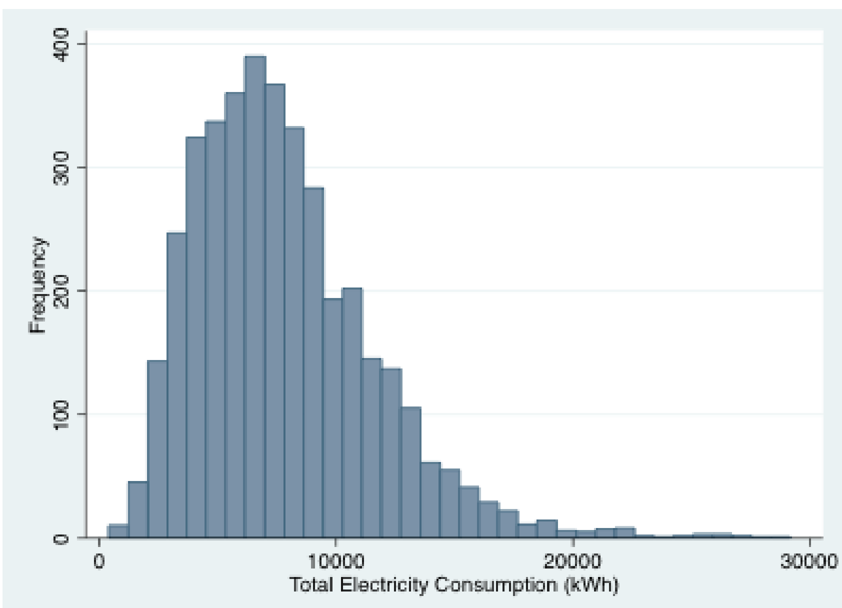

Figure 5 - Distribution of Electricity Consumption Among Non-Participants in the Top Ten Participant Segments in Eastern Sample

Modeling Participation Rates

Some of the socioeconomic indicators used to explain the difference in consumption variance can also be used to model program participation rates. Homogenous groups of older homes are significantly more likely to have high participation in energy efficiency programs in both samples than groups of newer homes. In the western sample, higher participation rates are found among homogenous groups of homes in relatively more urban areas, in neighborhoods with higher proportions of homeowners with older household heads. In the eastern sample, program penetration rates are positively correlated with proportion of college graduates in the neighborhood and average household size. While some of these variables were not included in the previously outlined models for explaining energy consumption variance, they are almost all collinear with variables in the variance models and affect consumption variation in the same directions. Key variables were left out of the penetration rate models due to co-linearity but nonetheless are important to note and are included in tables 21 and 23. The independent variables explain about 58% (.58 r-sqaured) of the variation in participant rates among homogenous groups of western homes and 85-88% of the participant variation in the eastern homes. This suggests that these variables can predict participation rates, in addition to energy consumption variation, fairly confidently.

Table 20: Variables That Explain Penetration Rate in Eastern Sample

Indicator	Relationship	T-Statistic	P-Value
Mean Year Built	Negative	-11.12	<.001
Mean HH Size	Positive	11.10	<.001
% College Education	Positive	35.12	<.001

Table 21: Variables Left Out of Eastern Penetration Rate Model due to Co-linearity

Indicator	Related To
% Single Family	Mean HH Size
% Home Owners	Mean HH size
% Income > $100k	% College Education
Median Householder Age	Insignificant
Home SQFT	Insignificant

Table 22: Variables That Explain Penetration Rate in Western Sample

Indicator	Relationship	T-Statistic	P-Value
Mean Year Built	Negative	-12.52	<.001
% Urban	Positive	2.02	.044
% Homeowners	Positive	2.79	.006
Median Age	Positive	6.85	<.001

Table 23: Variables Left Out of Western Penetration Rate Model due to Co-linearity

Indicator	Related To
% College Education	% Income > $100k
% Income > $100k	% Homeowners
Mean HH Size	Median Householder Age
Home SQFT	% Income > $100k
% Single Family	% Homeowners

The implications for targeting of households for energy efficiency measures from this paper are different for the two samples. We show that more program participation happens in areas of increasing differences in energy consumption among homogeneous households in both states. The program participants in the west are more diverse in their heating consumption than non-participants and their consumption tends to be on the higher end. The program participants in the east appear to be more homogenous in their total energy consumption and their consumption tends to be on the lower end. These are subtle tendencies which cannot be proven with statistical significance. What we observe in the big picture is a lot of diversity in consumption among homogenous households which also extends to groups of program participants.

This research applies a more rigorous and detailed account of customer targeting for energy efficiency measures. Customer targeting experience by the author suggests that households which have higher energy consumption, bigger homes, more income, are older and more educated are more likely to participate in energy efficiency programs. This research does not refute that established knowledge. However, when you force like households together and hold everything but energy consumption constant, what you see is that energy consumption varies significantly among those inclined to energy efficiency as well as those who are not. These findings suggest that marketing messaging in a “one size fits all approach” may not be the most effective. Different messages must be crafted to appeal to those who use a lot of energy as well those who already use less.

What This Research Means for Energy Efficiency Marketing

The diversity of energy consumption patterns and different household perspectives on energy consumption will require segmented marketing strategies to meet aggressive energy efficiency goals into the future. This topic deserves a whole paper itself but we will briefly touch on what this research means for marketing energy efficiency programs, and what we can learn from other business sectors that rely on data intensive segmented marketing.

The energy efficiency sector has done research on and used segmented marketing practices in recent years. Notable research includes studies by the Bonneville Power Administration[1], Ontario Power Authority[2], BC Hydro[3], and CIEE[4] and the author of this paper[5]. These studies indicate that attitudes towards energy and energy conservation vary significantly among households. These differing attitudes are due to much more than demographics. Attitudes are formed by world views and ingrained habits among other elements. This is why energy consumption varies dramatically even among homogeneous groups of homes whose households have similar demographics. We need to learn much more about how attitudes about energy use are formed and how those attitudes translate into different consumption patterns. With this information we can create a significantly more targeted approach to selling energy efficiency to a variety of customer segments.

The research in this paper verifies something that we already knew anecdotally, that consumption varies significantly in two very similar homes in the same place. This research has quantified how much we can expect this energy variability to be. Now we must set out to explain the variability. Some of the difference is due to the physical characteristics of the home, we believe that a lot of variability is due to unique behavioral patterns of people who live in homes. The methodology used in this paper to segment homes into homogeneous groups is a good starting place to develop more in depth studies involving targeted household interviews and surveys. We must know more about the small and large users of energy within homogenous housing segments. We need to know what causes their energy consumption patterns and what it will take each group to move forward with energy efficiency and conservation measures. The growing prevalence of AMI smart meters has the capability to greatly enhance our knowledge of when

1 Bonneville power Administration. "Residential Segmentation Research". March 2009. http://www.bpa.gov/Energy/n/segmentation/index.cfm

2 Ontario Power Authority. "Energy Conservation Attitudes and Behaviors (ECAB) Study" February 2012. http://www.powerauthority.on.ca/sites/default/files/2011_Energy_Conservation_Attitude_and_Behavior.pdf

3 Pederson, Marc. "Segmenting Residential Customers: Energy and Conservation Behaviors" 2008 ACEEE Summer Study on Energy Efficiency in Buildings. http://www.aceee.org/files/proceedings/2008/data/papers/7_671.pdf

4 Moss, Steven and Kerry Fleisher. "Market Segmentation and Energy Efficiency Program Design" Prepared for California Institute for Energy and Environment Behavior and Energy Program. November 2008. http://uc-ciee.org/downloads/MarketSegementationWhitePaper.pdf

5 Kelsven, Phillip. "Leveraging Data Mining and Geographic Information Systems to Gain Energy Efficiency Market Intelligence". ACEEE 2012 Summer Study in Buildings.

consumers are using energy and if combined with survey and interview data can produce powerful insights into how to move energy consumers to action. Using data and research in this manner is broadly referred to as "predictive analytics" and is being used successfully in other business sectors.

What We Can Learn From Other Business Sectors Using Segmented Marketing

Political campaigns have been on the cutting edge of segmentation research for many years. The age of big data has made their methods extremely effective and targeted. Much attention has been paid to President Obama's big data targeting strategy that helped him win the 2012 presidential election. The strategy involved merging information from various sources including voter records, fundraisers, consumer databases, field workers, and online social media sites. One article makes reference to seven different versions of a solicitation e-mail for a fundraiser[6]. The creators of Obama's segmentation and targeting machine have since gone on to much commercial success of their own consulting with eager businesses[7]

Retail businesses have increasingly used data and segmentation research to develop targeted marketing strategies in recent years. Most major retailers now have "predictive analytics" departments to make use of the trove of data they keep on customers. One company to successfully utilize these strategies is Target where they have marketed to distinct segments of their customer base. Target merges in store purchasing patterns with third party data on socioeconomic indicators to develop distinct customer segments that they use different marketing tactics and messaging on. Target has had particular success marketing to expecting mothers and families to get them to buy their baby products at Target. The company's algorithms are so good that they upset a father of a teenage daughter who was receiving baby product marketing materials. The father later found out that Target knew before he did that his daughter was pregnant[8].

Insurance companies have for a long time sought to know more about their customers behaviors. They want to know more about people's behaviors and habits not only to market to new customers, but to properly assess their own risk and the proper rate to charge a customer for providing an insurance policy. Progressive car insurance is using a product called "snapshot"[9]

---

6 "Crovitz, Gordon. "Obama's 'Big Data' Victory". Wall Street Journal. November 18, 2012. http://online.wsj.com/news/articles/SB10001424127887323353204578126671124151266

7 Rutenberg, Jim. "Data You Can Believe In". New York Times. June 20, 2013. http://www.nytimes.com/2013/06/23/magazine/the-obama-campaigns-digital-masterminds-cash-in.html

8 Duhigg, Charles. "How Companies Learn Your Secrets". New York Times Magazine. February 2012. http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=0

9 Stross, Randall. "So You're a Good Driver? Let's Go to the Monitor". New York Times. November 24, 2012. http://www.nytimes.com/2012/11/25/business/seeking-cheaper-insurance-drivers-accept-monitoring-devices.html

which tracks the time of day, distance travelled, as well as speed second by second in customers cars. Customers who volunteer for the program can receive up to a 30% discount on their car insurance. This is the kind of behavioral data that can be extremely valuable to a company like progressive who seeks to find out which behaviors make a safe driver and which make a risky driver. Traditionally, car insurance companies relied on age, sex, marital status, driving history, and credit scores to assess risk. There is no better data however than real time behavioral observation to assess risk in an insured driver. Over time Progressive can observe the outcomes (accidents or no accidents) of drivers using snapshot to assess risk and create better rates. This is the kind of device that makes researchers in behavioral energy savings jealous. Imagine what we could learn from a similar device that tracks energy use in homes. In some sense AMI smart meters do a similar job in utility customers homes by tracking data in up to 15 minute intervals, however it is still difficult to assess with accuracy what end uses customers are consuming energy on.

Healthcare is undergoing a revolution using data and predictive models. Nearly all medical records in the United States are now being collected electronically and many historical records are being converted to electronic records. This data is being used to improve health outcomes in people and reduce healthcare costs. Some healthcare systems are using data to predict which patients are at risk for particular ailments in the future[10] and then intervene successfully to prevent ailments. One hospital in Texas uses predictive models around intervention successes that has cut down on 30-day readmissions for Medicare patients with heart failure by 31%[11].

These stories all relate to using data on people and households to segment them into specific groups for marketing a product or getting them to take an action. The energy efficiency and conservation sector has a lot of data to gather before we can truly understand what drives people to save energy in their homes. We have a lot more than data to gather. We need to understand what emotional, ethical, and economical triggers cause people to invest in energy efficiency.

Conclusion

This study concludes that significant variation exists in energy consumption among two large samples of homogenous groups of homes . Variance measured in standard deviation and controlled for sample size is significant within segments which conservative estimates indicate have average spreads in consumption from factors of 2.5 to 5.8 for total energy consumption with higher spreads in electricity and gas consumption separately. The spread of energy consumption also appears to be greater in the eastern sample than the western sample. Homes in homogenous segments with the highest variations in energy consumption consume considerably more energy, on average, than segments with the lowest consumption variations. Energy

10 Jacobson, Gary. "How Big Data Can Change the Way Hospitals Treat Patients" Dallas News. May 25 2013. http://www.dallasnews.com/business/health-care/20130525-how-big-data-could-change-the-way-hospitals-treat-patients.ece?nclick_check=1

11 Jacob, Steve. "Health Systems Use 'Big Data' to Cut Costs, Improve Quality". Dallas – Fort Worth Healthcare Daily. June 13, 2013. http://healthcare.dmagazine.com/2013/06/13/health-systems-use-big-data-to-cut-costs-improve-quality/

consumption variance is also significantly different across as well as within homogenous housing segments.

The variation in standard deviations of energy consumption can be mostly explained by a number of socioeconomic factors, including the size and age of homes as well as income, age, and types of residential homes in the segments. There is more unexplained variation in consumption standard deviations in the western sample than the eastern sample.

Controlling for sample size only, the rate of residential participation in energy efficiency programs is positively related to variance in energy consumption. Participation rate can be largely explained by the same socioeconomic factors that explain differences in consumption variation. Median gas and total energy consumption and variances of gas and total energy consumption do differ somewhat among energy efficiency program participants and non-participants, however the differences are subtle and not statistically significant. The western sample participants tend to use more total energy than non-participants and the eastern sample participants tend to use less. The variances of total energy consumption do not differ among participants and among non-participants in the western sample and appear to differ somewhat in for the eastern sample. The gas consumption variance of western households is greater than that of non-participants, and in the eastern sample the variance of participants is lower among all energy types.

The implications for energy efficiency program administrators is that there is a lot of diversity in energy consumption among housing segments that we would expect to have similar energy consumption. The HVAC efficiency, appliance inventory, and electronics inventory as well as the occupant differences in behavior are major variables that could not be studied in this paper. All of these are causes of significant variations in energy consumption. A study methodology that uses the segmentation and clustering method in this paper combined with an inventory of home characteristics and associated energy using devices could more accurately estimate the behavioral energy component to the consumption differences documented in this paper.

It is clear that households use energy in significantly different ways. Households also invest in energy efficiency for many reasons. The case for targeting segments of customers along the whole energy consumption distribution differently may increase participation by customizing the message that households receive and drives them to action. It is time that our industry widely adopts segmenting and messaging practices arrived at from data collection and predictive analytics research which are successfully used in other industries.