Natural Gas Savings from Energy Efficiency Measures in the Home Energy Rebate Program: A Multivariate Regression Analysis

Dustin Madden, Cold Climate Housing Research Center *June 30, 2019*

Introduction

Savings for energy efficiency measures are typically based on modeled estimates of home energy consumption before and after implementing the retrofit. This study attempts to determine the actual natural gas consumption savings from implementing specific energy efficiency measures by using a multivariate linear regression analysis. Actual natural gas consumption data for a period of nine years was obtained from the Enstar natural gas utility and matched to AkWarm home energy rating software records for each home that participated in the Home Energy Rebate program. The goal of the regression analysis is to determine which energy efficiency measures saved the most energy in real-world conditions.

Methodology

There were three primary steps used to identify the natural gas consumption savings from individual energy efficiency measures in a statistically rigorous way:

1. Determine which energy efficiency measures were implemented for each household
2. Calculate the average annual natural gas savings per household using Enstar data
3. Conduct multivariate regression analyses

Determine which energy efficiency measures were implemented for each household

The first step was to determine which energy efficiency measures were implemented by comparing the pre- and post-rating files for each single family home that participated in the Home Energy Rebate program. Detailed data on each building shell component of each home was collected for both the pre- and post-rating. A Python script was developed to compare each of these components for changes in insulation levels and square footage for each home that participated in an energy efficiency retrofit. For insulation retrofits, a minimum threshold of an increase in r-value of at least 0.5 was implemented to filter out accidental changes caused by variation in the way energy raters entered the component versus changes caused by an actual retrofit to increase the insulation levels. Each of these retrofits was then additionally classified by its location within the house; for example, above-grade wall components were split into “garage walls”, “house walls”, “rim joists”, and “crawl space walls.” Additional analysis was conducted on heating, ventilation, water heating, air-tightness and control systems for each home to determine whether they were retrofit during the program and if so, in what way. Here are the key independent variables that were identified for use in the regression analysis:

- *Shell components:* Change in shell components for each category and location subcategory, expressed in terms of assembly delta u-value times the size of the shell component
    - Main categories: above grade floor, below grade floor, above grade wall, etc.
    - Location sub-categories: house, garage, crawl space, etc.
- *Ventilation systems:* dummy variables for installation of continuous mechanical ventilation or heat recovery ventilation systems
- *Air leakage*: calculated the change in modeled natural air changes per hour from pre- to post-rating and normalized by the volume of the home
- *Heating Systems:*
    - Split the heating systems into primary and secondary systems
    - Filtered out homes that did fuel switching
    - Calculated the change in AFUE
        - Used the percent of heating done by the primary and secondary systems (as entered by the rater through the radio buttons) to weight the change in AFUE
        - Weight the change in AFUE by the AkWarm modeled heat load
    - Zeroed out heating system changes for equipment not using natural gas (since it won't affect our dependent variable)
- *Domestic Hot Water systems:*
    - Filtered out homes that did fuel switching

- Zeroed out domestic hot water systems that didn't use gas since that wouldn't affect our dependent variable
- Calculated the change in energy factor

Calculate the average annual natural gas savings per household using Enstar data

Anonymous Enstar natural gas consumption data from 2009 to 2017 was used to calculate the average annual natural gas savings per household. The dataset included monthly gas consumption for each household as well as latitude and longitude. GIS techniques were used to match the Enstar data to each AkWarm energy rating that was available for homes participating in either the Home Energy Rebate or Weatherization programs. Essentially, each meter location was matched to a property tax parcel, which was then matched to the address from the AkWarm energy rating for each participating household. Approximately 82% of participating households were matched to a single Enstar meter location

After completing matching, the Enstar data was cleaned. Specifically, records that had less than zero consumption were removed from the analysis, and years of data that had fewer than a full twelve months were also filtered out.

To try to account for differences in heating requirements between years and locations the annual natural gas consumption was normalized by heating degree days. This was done by matching monthly heating degree data for each weather station in the region to the household and using it as the denominator to normalize the consumption. A heating degree day database was created using historical data obtained from a DegreeDay.net API, which uses only data from weather stations that have met quality control standards. The base-60 heating degree days for each month were matched to the Enstar natural gas consumption data using a zip code to weather station lookup table created through the same DegreeDay.net API.

Another method used in this analysis to try to account for the different weather over time and in different locations was to include the difference in heating degree days multiplied by the modeled heat load of the building as an independent variable. First, the average annual base-60 heating degree days was calculated for each weather station corresponding to each household that participated in a retrofit program. Then, the data was averaged for all years pre-energy rating and the difference between this and the average for all years post-energy rating was calculated. This difference in heating degree days was multiplied by the modeled AkWarm heating loads; the modeled heat load was used as a way to ensure that the heating degree days were only normalizing the requirements for heating, and not accidentally normalizing other natural gas end uses which are not necessarily affected by the climate, such as domestic water heating, cooking, and clothes drying.

Finally, the average annual natural gas consumption savings per household was calculated. The first step to this was to filter the dataset so that there was only data for complete years before the as-is energy rating date or after the post energy rating date. Only 48.5% of records had at least one full year of data before the as is rating *and* at least one full year of data after the post rating. The monthly gas consumption was summed for each complete year pre- and post-energy rating to create annual records for each year, and additionally an annual metric for BTUs per heating degree day was calculated by dividing by the base-60 heating degree day data. The average annual gas consumption as well as the average annual BTUs per heating degree day for all years before the as-is rating and all years after the post rating was calculated for each participating household. Using this final filtered dataset of 6,544 matched as-is and post rating pairs, the average number of years before the as-is rating was 2.9 and the average number of years after the post rating was 3.8. The very last step was to calculate the average annual natural gas consumption savings and the average annual BTU per heating degree day savings for each participating household to be used as the dependent variable for the multivariate regression analysis.

Additionally, in order to try and account for any possible variation potentially caused by different sets of occupants living in the homes that were retrofit, the annual natural gas consumption savings was also calculated using just a single complete year before the as-is energy rating and a single complete year after the post energy rating. Other research has shown occupant behavior to be a potentially large factor in the energy consumption of residential buildings; while the single-year method does not completely rule out the possibility of different occupants, by shortening the time frame it reduces the likelihood that there would have been a different set of occupants living in the home than those who participated in the energy efficiency retrofit program.

Conduct multivariate regression analyses

Multivariate regression analyses were conducted using three different dependent variables: the annual natural gas consumption savings, the annual BTU per heating degree day savings (i.e. the climate-normalized annual natural gas consumption savings), and the single-year annual natural gas consumption savings. This was done using all of the independent variables detailed in the methodology. Another set of regressions were run using each independent variable on its own to determine which of them were statistically significant and then the subset of the data that only includes all of the variables identified as significant was used in a final regression.

Results

Overall, the regressions produced relatively low r-squared values, ranging from 0.19 to 0.26. This suggests the independent variables only account for a small portion of the overall variance in savings between homes. Table 1 highlights the results from the different permutations of regressions that were run.

Table 1: R-squared values for different multivariate regressions (N=3,410)

Dependent variable	R-squared value - all variables	R-squared value - statistically significant variables only
Average annual natural gas savings	0.211	0.204
Single-year annual natural gas savings	0.263	0.257
Average annual BTU per heating degree day savings	0.199	0.194
Average annual natural gas savings with heating degree days * heat load	0.215	0.209

It is likely that the individual year variation in natural gas consumption and climate decreases the likelihood of a strong correlation in terms of r-squared value; the higher r-squared value for the single-year annual natural gas savings as compared to the average savings from pre- to post-energy retrofit provides evidence that this is the case.

There were several energy efficiency measures that consistently were highly statistically significant ($p < 0.01$) and which had relatively tight 95% confidence intervals. The four primary energy efficiency measures that have proven savings were upgrading the efficiency of the primary heating system, increasing the efficiency of the water heating system, increasing the air-tightness of the home, and insulating the ceiling. The confidence intervals for these variables and the interpretation of the projected savings are detailed in Table 2.

Table 2: Confidence intervals and interpretations for energy efficiency measures with strong statistical significance (N=3,410)

Energy Efficiency Measure	EEM installation counts	Average change in component efficiency	Efficiency units	low avg gas savings estimate (ccf/year)	projected avg gas savings estimate (ccf/year)	high avg gas savings estimate (ccf/year)
Insulate ceilings	1,353	0.03 * 1,089 sqft	U-value * square footage	25	49	73
Increase air-tightness	3,086	2809.00	Natural ACH x volume	76	100	124
Upgrade water heater	2,406	28.70	Energy factor	119	143	167
Upgrade heating system	2,932	19.33	AFUE	94	115	137

Discussion

The generally low r-squared values suggest a lot of unexplained variance in the data. There are a variety of factors that could be contributing to this, including:

- **Occupant behavior:** Occupant behavior is widely recognized as a significant factor contributing to variation in home energy consumption.[1] Differences in the use of space heating equipment and maintenance, temperature setpoints, domestic hot water consumption, operation of doors and windows, and appliance usage can all contribute significantly to the energy requirements of a home. These factors are not accounted for in this study, and likely contribute to the unexplained variance in the natural gas consumption over time, especially since the study period covers multiple years.
- **Collinear variables:** Savings from heating systems and increasing the efficiency of the building through air-tightening and insulating will affect each other, as decreasing the

[1] Yan, D., O'Brien, W. et. al. *Occupant behavior modeling for building performance simulation: Current state and future challenges.* (2015). Energy and Buildings. Available at: https://escholarship.org/dist/wJD5zTgC2vrImRR/dist/prd/content/qt8755h7rn/qt8755h7rn.pdf

heating load will reduce the total natural gas savings from increasing the efficiency of a heating system by a given amount and vice versa. These interactions between energy efficiency measures mean that the variables will have some level of collinearity, decreasing the statistical power of regression analysis. Additionally, some building energy efficiency retrofits necessarily will almost always affect multiple components. For example, adding exterior foam insulation to walls will also increase the air-tightness of the building, and installation of a new heating system will likely also include a programmable setback thermostat. In these and many other cases these individual variables included in the analysis will be collinear, as they are correlated with each other in addition to being correlated with the dependent variable natural gas consumption.

- **Weather variation:** While the regression analyses conducted in this study tried to account for variation in the weather by using heating degree day data from the nearest weather station, some of the unexplained variance in the data is likely due to differences in microclimates and non-temperature related factors. Even with the study area being limited to homes in the Enstar region, weather can vary significantly from place to place, and Alaska has a limited set of weather stations that have adequate data quality and quantity. Additionally, other factors such as wind or the amount of solar irradiation reaching a home can affect energy consumption and were not considered in this analysis.

- **Temporal changes in occupancy / building characteristics:** Finally, over the course of the eight years of the study period it is possible that there were changes in occupancy or additional changes to building characteristics that would not be accounted for in this analysis. Different occupants may have significantly different behaviors affecting energy consumption. Additionally, if there were changes to the home beyond those documented between the pre- and post-energy rating, they would not be accounted for in this study, and could potentially affect energy consumption in either direction. Both of these factors likely contribute to some of the unexplained variance in the dataset.

There were four energy efficiency measures that were consistently statistically significant to a p-value of less than 0.01 and had relatively tight confidence intervals. While the regression analyses suggest that these measures reliably save energy and provide a confidence interval with real-world natural gas consumption savings, it does not necessarily mean that other energy efficiency measures do not provide savings. The same factors that contribute to the relatively low r-squared values in this study also contribute to lower measures of statistical significance for independent variables; additionally, many of the energy efficiency measures being analyzed were only implemented in a small number of homes in this dataset, which decreases the likelihood of them returning significant results in the regression analyses.

Recommendations

There is a high percentage of unexplained variance in the data, despite having access to very detailed records of the physical changes made to building characteristics for households participating in the Home Energy Rebate program. Of the factors identified in this study, occupant behavior variation is the most likely contributor to this variance in the data. Previous energy efficiency programs in Alaska have focused primarily on building energy retrofits, which have successfully reduced energy costs and consumption, stimulated the economy, and generated local jobs. If indeed occupant behavior accounts for a significant amount of the variance in energy consumption between households, then it means there is an untapped source of potential future energy savings through behavior-based programs. Given this potential for savings and the surge of recent research on the topic of occupant behavior, we recommend conducting a detailed review of the scientific literature on the effects of occupant behavior on energy consumption in cold climates and the effectiveness of behavior change programs. If the evidence supports the effectiveness of such programs, this would represent a relatively untapped opportunity to reduce energy costs and consumption in Alaska.