Installed Cost of Energy Efficiency Measures in the Home Energy Rebate Program: A Multivariate Regression Analysis

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

Introduction

A key piece of information needed to determine whether or not a building energy efficiency measure is cost-effective is its actual installed cost. While there are national estimates for a variety of energy efficiency measures[1], the costs vary both temporally and regionally, making accurate estimates for a specific time and place difficult to obtain. This study attempts to use a multivariate regression analysis to determine the installed costs of energy efficiency measures that were implemented between 2008 and 2018 as a part of the Alaska Housing Finance Corporation's (AHFC) Home Energy Rebate Program.

While there are a number of end uses for the cost data for energy efficiency retrofits, one key place it is currently used is in the AkWarm Home Energy Rating software. AkWarm is used by Energy Raters and Weatherization Assessors statewide as a tool to determine which energy efficiency measures to implement. The software uses cost data to calculate cost benefit ratios for every combination of energy efficiency retrofits suggested by the rater and creates a list showing energy efficiency measures ranked from the most cost effective to the least. Ensuring that the cost data is accurate and current would lead to better investments and increased savings for each home that is retrofit.

Methodology

There were three primary steps used to try to identify the installed costs of energy efficiency measures in a statistically rigorous way:

1. Determine which energy efficiency measures were implemented for each household

[1] The most complete of these is the National Residential Efficiency Measures Database. These costs are aggregated from performance home contractors, RS Means, the California Database for Energy Efficient Resources, etc. and analyzed and cleaned statistically.

2. Obtain and match reported total costs for each home to implement the energy efficiency measures

3. Conduct a variety of 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
    - Calculated the change in AFUE
- *Domestic Hot Water systems:* change in energy factor

Obtain and match reported total costs for each home to implement the energy efficiency measures

AHFC has data on the total costs homeowners paid for energy efficiency measures that they implemented through the Home Energy Rebate program. These costs are based on receipts submitted to AHFC, but are not itemized in the database. An SQL query was used to obtain data on the total owner costs per home, as well as the total rebate amount paid to the homeowner by AHFC and the increase in Star Steps from the pre- to post-rating. The total cost per home was used as the dependent variable in one set of regressions.

In addition to the total costs per home for the full dataset, a smaller subset of the data was created by identifying all of the participating homeowners who accomplished their retrofit for less than the rebate they were eligible for based on their improvement in energy efficiency. This was done by calculating the rebate they should have earned based on their increase in building efficiency and comparing that to the actual rebate they received and the reported receipts. This data subset was created in response to anecdotal evidence that homeowners did not submit all their receipts when they spent more on energy efficiency retrofit work than they were eligible to be reimbursed for. This limited dataset should account for this potential issue, using the assumption that if a homeowner was eligible to receive a larger rebate based on the level of efficiency they achieved, they would have submitted all of their receipts as every dollar spent would be a dollar they would earn.

After obtaining the costs of implementing energy efficiency measures for each household participating in the Home Energy Rebate program, costs were then adjusted for inflation. Households paid for the retrofits over a period from 2008 to 2018, and so to make them comparable we used the "Urban Alaska" consumer price index data from the Alaska Department of Labor and Workforce Development.[2] While this does not take into account year-to-year market fluctuations within the construction materials industries, it does account for the overall changes in the total market.

Conduct a variety of multivariate regression analyses

Multivariate regression analyses were conducted on a variety of subsets of data and dependent variables using the ordinary least squares method. One set of regressions was conducted using the inflation-adjusted homeowner paid costs as the dependent variable, and another set was done on the unadjusted homeowner paid costs with the dependent variable along with an additional independent variable of the year in which the homeowner paid for the energy efficiency measures.

[2] Available at: http://live.laborstats.alaska.gov/cpi/index.cfm

Regressions were conducted for both sets of dependent variables using a variety of subsets of the data, including the following:

- All homeowners participating in the Rebate Program
- Only homeowners that received less than the rebate that they qualified for
- Datasets for participants from each of the following regions:
    - Municipality of Anchorage
    - Matanuska-Susitna and Kenai Peninsula Boroughs
    - Fairbanks North Star Borough
    - Sealaska ANCSA Region
    - All other regions together, called “Rural Areas” for this study

Finally, regressions were done on all of the independent variables described above, as well as a subset of the variables that only included those that were statistically significant.

Results

All Regions

Despite the large dataset, as a whole, the r-squared values were fairly low for all multivariate regression analyses. The highest statewide r-squared value was 0.187, which means that more than 80% of the variance in the costs is not accounted for by the independent variables used in the regression. Of the two datasets used, it appears that the smaller subset that only included homeowners that accomplished their energy efficiency measures for less than the rebate they were eligible for had a stronger correlation, as can be seen in Table 1.

Table 1: R-squared values for statewide multivariate regression analyses

	Inflation Adjusted, all variables	Not Inflation Adjusted, all variables	Inflation adjusted, stat significant variables only	Not inflation adjusted, stat significant variables only
All homeowners	$R^2 = 0.111$ n = 18,789	$R^2 = 0.113$ n = 18,789	$R^2 = 0.109$ n = 18,789	$R^2 = 0.111$ n = 18,789
Homeowners receiving less than their qualified rebate	$R^2 = 0.175$ n = 3,306	$R^2 = 0.187$ n = 3,306	$R^2 = 0.170$ n = 3,306	$R^2 = 0.185$ n = 3,306

The 95% confidence intervals for the independent variables used in this analysis generally have fairly large ranges, even when their p-values show a very strong statistical significance. Table 2 shows a select set of confidence intervals for only the variables that are statistically significant (i.e. have a p-value of less than 0.01) and where the top end of the confidence interval is around 60% higher than the low end or less. The table highlights the coefficients from the regression done on all homes using inflation-adjusted homeowner costs. In addition, the table includes the interpreted total costs of the energy efficiency measures, which were estimated using the average sizes and changes in efficiency of each measure.

Table 2: Confidence intervals and interpretations for energy efficiency measures with strong statistical significance (N=18,789)

Energy Efficiency Measure	Low confidence interval coefficient	High confidence interval coefficient	average area	average change in component efficiency (u-value, energy factor, natural ACH*volume)	Typical EEM Cost - Low	Typical EEM Cost - High
Insulate above grade walls	37.00	51.00	782	0.06	$1,765	$2,433
Insulate ceilings	18.00	28.00	1,058	0.03	$627	$976
Replace windows	71.00	87.00	86	0.45	$2,754	$3,375
Replace water heater	7,563.23	9,077.49		0.25	$1,906	$2,288
Increase air-tightness	0.21	0.29		2,302.00	$483	$668
Install heat recovery ventilation	2,394.00	3,873.00			$2,394	$3,873

Regional Analysis

On a regional basis, the levels of correlation between the energy efficiency measures implemented and the reported costs varied significantly. There are a variety of factors that may have contributed to this, including differing levels of data (131 households at the lowest and over 11,000 at the highest), potential cost differences within regions, and many more. For all regions, there was a significant increase in the r-square value for the subset of data that included only homeowners that did not receive the rebate that they were qualified for based on their efficiency gains. This suggests that there is some validity to the anecdotal evidence that homeowners did not submit all of their receipts beyond what they needed to turn in to receive their rebate, as a lower variance when this was accounted for implies that whether or not all receipts were turned in was a factor.

Table 3: R-squared values for regional multivariate regressions using inflation adjusted costs and all independent variables

	Anchorage	Fairbanks North Star Borough	Mat-Su and Kenai Peninsula Boroughs	Rural Areas	Sealaska Region
All homeowners	*R^2 = 0.152* *n = 11,552*	R^2 = 0.144 n = 2,192	R^2 = 0.113 n = 2,855	R^2 = 0.176 n = 564	R^2 = 0.073 n = 1,626
Homeowners receiving less than their qualified rebate	*R^2 = 0.271* *n = 1,779*	R^2 = 0.194 n = 324	R^2 = 0.311 n = 687	R^2 = 0.341 n = 131	R^2 = 0.150 n = 385

A regional analysis of the confidence intervals generally showed wider ranges and less statistical significance, which is likely due to the lower sample sizes that come with using only subsets of the data. The one exception is the Anchorage region, which showed confidence intervals and significance levels that were very similar to the statewide results; this is likely due to the fact that most of the homes with sufficient data used in this study come from the Anchorage region.

Discussion

While there are some energy efficiency measures that were included in the regression analyses that have a reasonably small 95% confidence interval for the estimated cost of installation, overall the low r-squared values and generally large confidence intervals point to multivariate regression being an inadequate method to extract individual measure cost data from this particular dataset. There are a variety of factors that likely contribute to the results generally having low statistical validity, including:

- **Price variability between contractors:** There are likely significant differences in the price of installed energy efficiency measures between different contractors even in the same region of the state.
- **Regional price variation:** Many of the regions are large and could have significant variation even within a single region. The size of the dataset is likely not large enough to support more analysis of more granular regions.

- **Contracted versus personal labor:** Homeowners had the option of performing the energy efficiency retrofits in a manner of their choice; thus some will have performed all the labor themselves and just used the rebate to pay for materials, others will have contracted the labor out, and there will be a variety of combinations of both included in this dataset.

- **Collinear variables:** 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 of total retrofit cost.

- **Partially accounted for factors:** There are some factors that may have only been partially accounted for in this analysis. Examples include:

    - *Receipts:* The issue of incomplete submittal of receipts is likely still affecting the analysis, as not receiving the full rebate they were eligible for does not necessarily mean that those homeowners actually submitted all of their receipts. Additionally, reducing the data to this subset drastically decreases the sample sizes, which negatively impacts the statistical validity of the analysis.

    - *Temporal price variation:* While this analysis tried to account for time-based cost variation by using the year as an independent variable or by adjusting the costs using the historical change in the consumer price index in Alaska, there are several reasons this may be inadequate. One, the analysis did not account for potential seasonal cost variations due to contractor workloads varying between summer and winter or to added costs to perform retrofits in times of difficult weather conditions. The cost of materials and labor also may not have tracked the consumer price index used to try to normalize the prices to a base year. Finally, the identified year in which the costs were paid may have been incorrect by +/- 1 year due to a lag between the time the homeowner paid for the services and the time that the rebate was processed.

Recommendations

There is definite value to having current, accurate cost data on energy efficiency measures in Alaska: it would make software recommendations on which energy efficiency measures to implement more accurate, would increase trust in the reliability of energy efficiency to provide economic benefits to occupants, and would allow designers to determine which

energy efficiency measures should be included in new construction. However, due to the variation in the data and the lack of statistical validity using multivariate regression to try to identify the costs for the majority of the energy efficiency measures implemented in the Home Energy Rebate program, we recommend that additional methods be explored to track and report this data. Based off of the limitations discovered in this study, the following key characteristics need to be tracked in order to obtain accurate cost data:

- Costs need to be split into labor versus materials for each energy efficiency measure
- For shell components, it is important to track the area that is being improved, the difference in insulation value, and to properly categorize the component in terms of type and location
- Costs for individual measures need to be tracked rather than total cost of the retrofit
- The year in which the measure was installed