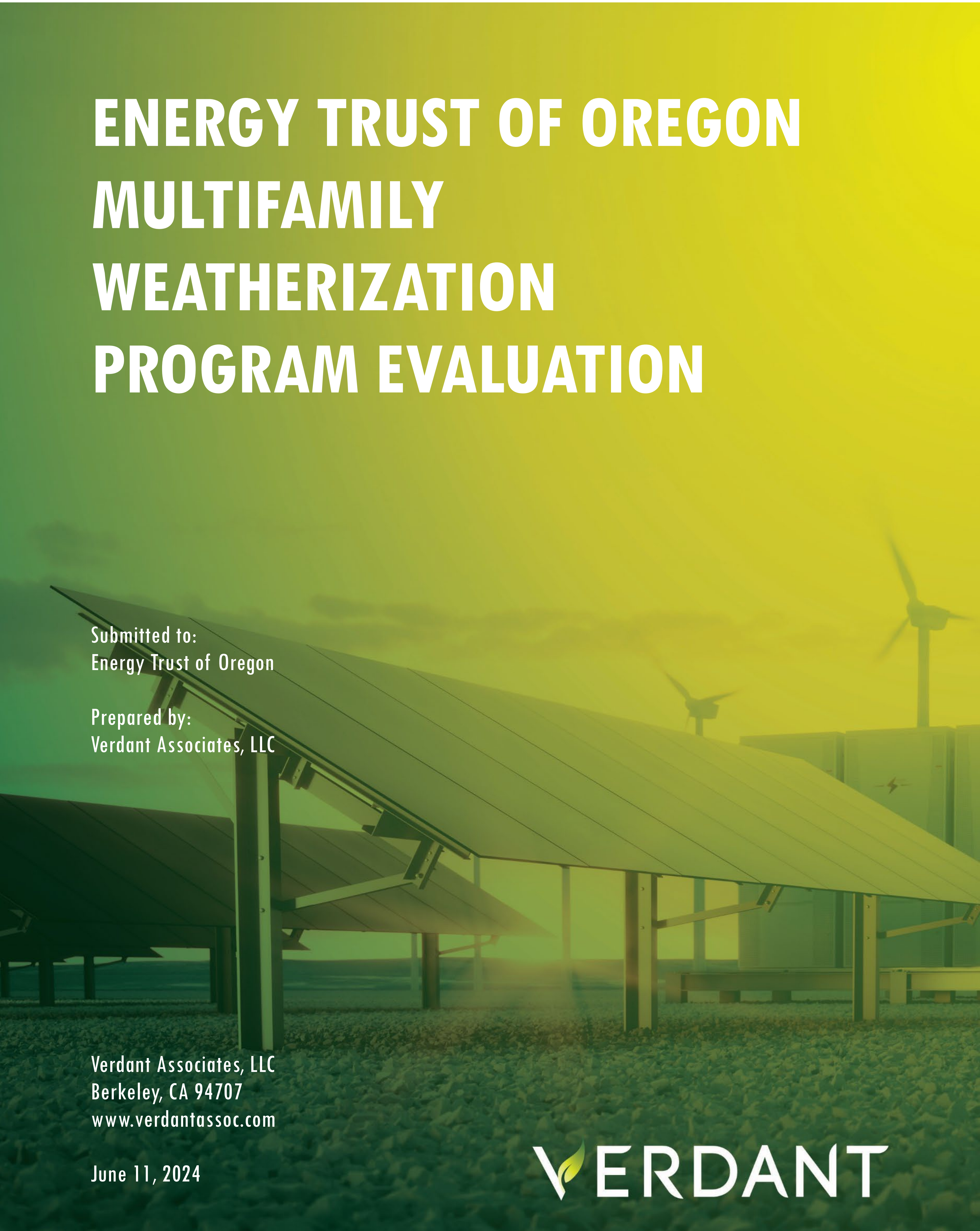
ENERGY TRUST OF OREGON MULTIFAMILY WEATHERIZATION PROGRAM EVALUATION



Submitted to:
Energy Trust of Oregon

Prepared by:
Verdant Associates, LLC

Verdant Associates, LLC
Berkeley, CA 94707
www.verdantassoc.com

June 11, 2024

VERDANT

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

VERDANT

1 EXECUTIVE SUMMARY

Energy Trust of Oregon provides cash incentives for multifamily properties to make energy-related improvements to the building shell. Incentivized improvements include window retrofits and upgrades, and insulation retrofits. This evaluation assesses the energy savings from installations completed in program years 2016 to 2021. Table 1-1 presents the savings claimed by the program between 2016 and 2021.

TABLE 1-1: MULTIFAMILY WEATHERIZATION CLAIMED SAVINGS & PROJECTS BY MEASURE AND HEATING TYPE

Project Measure	Building Size	Claimed kWh	Claimed Therms	Claimed SQFT	Number of Projects
Electric Heat					
Window Upgrades	Small	967,734	0	237,580	922
Window Retrofit	Large	5,096,756	0	741,817	488
Insulation	All	974,347	0	1,513,104	476
Gas Heat					
Window Upgrades	Small	3,879	39,697	132,086	579
Insulation	All	26,131	17,427	356,498	217

The primary objective of this evaluation is to assess the program's electric and natural gas savings from incentivized measure installations by various domains of interest. Savings are categorized by measure type (window retrofits, window upgrades, and insulation), heating type (electric and gas), and building size (small and large).[1] Claimed savings can vary between small and large buildings due to differences in assumed baselines or building thermodynamics and their modeled energy usage.

Table 1-2 presents the estimated average energy savings per square foot of installed measures and the realization rates. The observed realization rates compare the modeled savings to the ex ante values. The ex ante values for window upgrades, however, represent market baselines while the estimated savings are relative to existing conditions. For windows upgrade measures installed at small sites, we calculated adjusted realization rates that reflect the estimated savings relative to an adjusted ex ante existing conditions baseline. The savings and realization rates are developed using data from program years 2016 to 2021.

The evaluation found that the realization rate for window retrofits and insulation in electric heated multifamily buildings is 65 and 67 percent, respectively, of ex ante values while the realization rate for insulation in gas heated buildings is 90 percent of the ex ante values. For window retrofit and insulation

1 Small multifamily is two to four dwelling units or a side-by-side configuration and large multifamily is five or more dwelling units in a stacked configuration.

measures, the estimated realization rates are less than 100 percent, but the confidence intervals include a 100 percent realization rate, implying that the ex ante savings for these measures are not inconsistent with the study's measured values. Window upgrades are estimated to provide higher savings per square foot of window area than window retrofits but the estimated adjusted realization rate for window upgrades is only 34 percent for electric heated buildings and 10 percent for gas. For window upgrades, the adjusted realization rate and estimated confidence interval, is less than 100 percent, implying that the study's estimates of saving are statistically lower than the ex ante values.

TABLE 1-2: ESTIMATED SAVINGS SEGMENTED BY MEASURE, HEATING TYPE, AND SIZE

Energy	Measure	Building Size	Model n***	Estimated Energy Savings per SQFT Installed**	Realization Rate Observed	Realization Rate Adjusted
Electric Heat						
kWh	Window Upgrade	Small	320	5.61 ± 2.08	133% ± 49%	34% ± 12%
kWh	Window Retrofit	Large	100	4.87 ± 3.14	65% ± 42%	
kWh	Insulation	Small	140	0.45 ± 0.42	67% ± 63%	
kWh	Insulation	Large	0	Not evaluated	Not evaluated	
Gas Heat						
therm	Window Upgrade	Small	229	0.12 ± 0.09	40% ± 30%	10% ± 8%
therm	Insulation	Small	35	0.04 ± 0.02	90% ± 47%	
therm	Insulation	Large	0	Not evaluated	Not evaluated	

**For this and other columns with the "±" following the value, the second value indicates the 90 percent confidence interval.

***The "Model n" is the total number of install periods for each Analysis ID that went into the model. See section 3.2.2 for a description of Analysis ID.

The results presented in Table 1-2 were estimated using data for structures with nine or fewer units participating in the program. These results may not be representative when applied to the participant population due to the small share of Analysis IDs whose data were included in the model.[2] Due to data attrition issues, the study was not able to assess the savings for measures installed in campuses or structures with 10 or more units. These installations represent the minority of structures with installed measures but account for the majority of the program's savings. To remedy this in future program years we recommend the program implementer collect the following data in the program tracking database:

- A record for each individually treated unit, including address, apartment number, the square footage of insulation or windows installed in the unit, the premise id and meter number (linking directly to the billing data), an indicator to note if the unit is master-metered, the floor(s) the unit is on, and an identifier to associate other units that are in the same building.

[2] For electric heat measures the model includes approximately 30 percent of the original Analysis IDs with a higher share for small windows and insulation than larger window retrofits. The gas models include approximately 35 percent of the original Analysis IDs.

- For insulation measures that may affect multiple units, collect information on each unit the insulation affects and an identifier to associate units that are in the same building.
- For complexes and structures that are rentals, collect an average vacancy rate from the building owner as part of the application process.

EnergyTrust of Oregon

MEMO

Date: 7/22/2024
To: Energy Trust Board of Directors
From: Dan Rubado, Sr. Project Manager – Evaluation
Sarah Castor, Evaluation and Engineering Manager
Patrick Urain, Sr. Program Manager – Commercial
Subject: Staff Response to the 2016-2021 Multifamily Weatherization Evaluation

Verdant Associates conducted a billing analysis of Energy Trust funded weatherization measures installed in multifamily buildings from 2016 to 2021, including insulation and efficient windows. This analysis was beset by data quality issues, particularly with larger buildings, so the results represent only a small portion of the program's total claimed energy savings for these measures. Because of these problems, the results have relatively low precision and can only be applied to multifamily buildings with fewer than 10 units. In addition, there may be comparability issues when trying to generalize the results to the broader program population, even among smaller buildings. Results from this study will be considered as one of many inputs into Energy Trust's measure development process to update deemed savings values for multifamily windows and insulation measures.

Verdant found window retrofits in buildings with five to nine units had electric savings of 4.87 kWh per square foot installed and a realization rate of 65%. However, due to low precision, this was not statistically different from the savings claimed by the program. Window upgrades measures in small buildings with two to four units had very low realization rates, at 34% for electricity and 10% for gas, both significantly lower than the savings claimed by the program. Alternatively, when the window upgrade measures were analyzed as retrofits, then savings were much higher—5.61 kWh and 0.12 therms per square foot installed for electric- and gas-heated buildings, respectively. This information may be helpful in assessing the cost-effectiveness of efficient window retrofits in small multifamily buildings, for which Energy Trust does not currently provide incentives.

Verdant found insulation measures in buildings with two to four units had electric savings of 0.45 kWh and gas savings of 0.04 therms per square foot installed, with realization rates of 67% and 90%, respectively. Due to low precision, the evaluated savings for these measures were not statistically different from the savings claimed by the program. However, given the low electric realization rate, it is likely the true average electricity savings per square foot of insulation in small multifamily are lower than claimed.

Verdant identified several significant data quality issues that complicated its analysis and limited the applicability of the results. They also provided a number of recommendations for improving Energy Trust's data entry processes and data systems to enable better analysis of multifamily buildings in the future. To address these data issues, Energy Trust will take the actions listed below, based on Verdant's recommendations.

First, many multifamily weatherization measures were installed only in a subset of dwelling units, but the individual units were not recorded in Energy Trust Project Tracking data, which made it more difficult to evaluate the impacts. To address this issue and improve future multifamily evaluations, Energy Trust's Existing Buildings program will begin recording the following information during multifamily project data entry and incentive processing:

- Identify individual dwelling units that receive incentivized measures in Energy Trust's Project Tracking system and link them to distinct building and dwelling unit identifiers in Energy Trust's CRM and Utility Customer Information systems.
- Record the address and unit number, building and unit site identifiers, utility meter and premise identifiers, quantity of measure installed in the unit, size of the unit, floor level, and a flag for master-metered buildings.
- Clean up multifamily site hierarchy information in Energy Trust's CRM system to identify distinct buildings on a campus and which units are located in each building.

Second, Energy Trust plans to begin recording the average vacancy rate for multifamily buildings installing weatherization measures. This information will be used to adjust the level of savings claimed for prescriptive measures and assess overall vacancy rates. Incorporating more accurate vacancy rates will align deemed savings values more closely with billing analysis results.

Third, Energy Trust's evaluation team will use alternative methods to validate measure assumptions and energy savings claims in multifamily buildings in the future, particularly in large multifamily buildings. These methods may include surveys, site visits and metering studies. This type of intensive evaluation work will be conducted as part of program-wide impact evaluations or as stand-alone research.

2 INTRODUCTION

2.1 DESCRIPTION OF THE PROGRAM

Energy Trust of Oregon provides cash incentives for multifamily properties to make energy-related improvements, including to the shell of their buildings. Building shell improvements include window retrofits and upgrades, and insulation retrofits. Multifamily properties eligible for the program include duplexes, triplexes, fourplexes, side-by-side units and stacked structures with five or more units. These multifamily properties must have heating provided by Portland General Electric, Pacific Power, NW Natural, Cascade Natural Gas, or Avista.

For the program years from 2016 to 2020, the cash incentives for multifamily shell improvements were provided by a standalone Multifamily program, implemented by Lockheed Martin. In 2021, Energy Trust's Existing Buildings program, implemented by TRC Companies, took over services to the multifamily sector.

Energy Trust's multifamily services have included both small and large multifamily buildings, including duplexes and townhouses all the way to large apartment and condo towers. Savings for small multifamily efficiency measures are developed separately from measures for larger buildings, using the same assumptions and savings calculations as single-family residential structures (since they share many characteristics). Measures for large multifamily buildings use different assumptions and calculations based on the characteristics and thermodynamics of larger structures with stacked units. This evaluation assesses the energy savings from installations completed in program years 2016 to 2021.

2.2 CLAIMED SAVINGS

Between 2016 and 2021 Energy Trust incentivized 2,463 multifamily weatherization projects that claimed 7,068,846 kWh savings and 57,125 therms savings. This section provides the claimed savings broken down by measure, heating type, building size, region, and market type.

2.2.1 Claimed Savings by Program Measure

Table 2-1 lists the multifamily weatherization savings claimed by Energy Trust for windows and insulation broken down by type of measure. The Program's window installations are disaggregated into window upgrades and retrofits. The window upgrades occur in either gas or electric heated small buildings while the window retrofits are exclusively installed in larger buildings heated with electricity. The insulation claimed savings are broken down by ceiling, floor, wall, and knee wall installations.

TABLE 2-1: MULTIFAMILY WEATHERIZATION CLAIMED SAVINGS AND PROJECTS BY MEASURE

Project Measure	Claimed kWh	Claimed Therms	Claimed SQFT	Number of Projects	kWh per Measure SQFT**	Therms per Measure SQFT	SQFT per Project
				Windows			
Windows Overall	6,068,369	39,697	1,111,483	1,977	5.46	0.04	562
Upgrades	971,613	39,697	369,666	1,490	2.63	0.11	248
Retrofit	5,096,756	0	741,817	488	6.87	0	1,520
				Insulation			
Insulation Overall	1,000,478	17,427	1,869,602	686	0.54	0.01	2,725
Ceiling	702,541	13,307	1,301,840	399	0.54	0.01	3,263
Floor	238,040	2,473	476,382	208	0.50	0.01	2,290
Wall	58,772	1,607	88,407	68	0.66	0.02	1,300
Knee Wall	1,125	40	2,974	11	0.38	0.01	270

**The window claimed kWh savings per square foot are relative to market baselines for window upgrades and relative to existing conditions for window retrofits contributing to higher claimed savings per square foot for retrofit windows.

2.2.2 Claimed Savings by Heating Type

A summary of the windows and insulation projects and their claimed electric and natural gas savings is presented in Table 2-2. These data provide an overview of the types of savings claimed by the program during the evaluation period (2016-2021). Window installations led to substantially more projects and claimed electricity savings than insulation projects. The per square foot savings for these projects are also presented, showing that the savings per square foot are higher for windows than insulation.

TABLE 2-2: MULTIFAMILY WEATHERIZATION CLAIMED SAVINGS AND PROJECTS BY MEASURE AND HEATING TYPE

Project Measure	Claimed kWh	Claimed Therms	Claimed SQFT	Number of Projects	kWh per Measure SQFT	Therms per Measure SQFT	SQFT per Project
				Electric Heat			
Window Upgrades	967,734	0	237,580	922	4.07	0	258
Window Retrofit	5,096,756	0	741,817	488	6.87	0	1,520
Insulation	974,347	0	1,513,104	476	0.64	0	3,179
				Gas Heat			
Window Upgrades	3,879	39,697	132,086	579	0.03	0.30	228
Insulation	26,131	17,427	356,498	217	0.07	0.05	1,643

The data presented in Table 2-2 indicate that sites with electric heating are responsible for the large majority of window and insulation installations and average substantially higher square footage of installed measures per project than gas heated sites.[3] The program tracking data for gas heated installations often show lower or zero electric savings, likely associated with air conditioning and/or furnace fan savings. Note, some projects claimed both gas and electric heat, so the gas and electric rows do not sum up to match the totals presented above.

2.2.3 Claimed Savings by Building Size

Energy Trust maintains two building size categories of multifamily efficiency measures:

- Small multifamily: two to four dwelling units or a side-by-side configuration
- Large multifamily: five or more dwelling units in a stacked configuration

The efficiency measures are developed separately by building size, using different assumptions and calculations. Claimed savings varies between small and large buildings due to differences in the thermodynamics of the buildings, and therefore their energy usage. Table 2-3 presents the program tracking savings by building size. Over 75 percent of the window projects and over 90 percent of the insulation projects are installed in small buildings. The program measures also claim a vast majority of the overall natural gas savings in small buildings. The average installation size of small building projects, however, is substantially smaller than projects installed in larger buildings. The project size differential, along with the claimed kWh savings per square foot of installed window,[4] contributes to the tracking data showing that windows in large buildings account for almost 80 percent of the claimed electric window savings.

TABLE 2-3: MULTIFAMILY WEATHERIZATION CLAIMED SAVINGS BY BUILDING SIZE

Project Measure	Claimed kWh	Claimed Therms	Claimed SQFT	Number of Projects	kWh per Measure SQFT	Therms per Measure SQFT	SQFT per Project
Small							
Window Upgrades	971,613	39,697	369,666	1,490	2.63	0.11	248
Insulation	727,029	16,908	1,250,604	636	0.58	0.01	1,966
Large							
Window Retrofit	5,096,756	0	741,817	488	6.87	0	1,520
Insulation	273,449	519	618,998	51	0.44	0	12,137

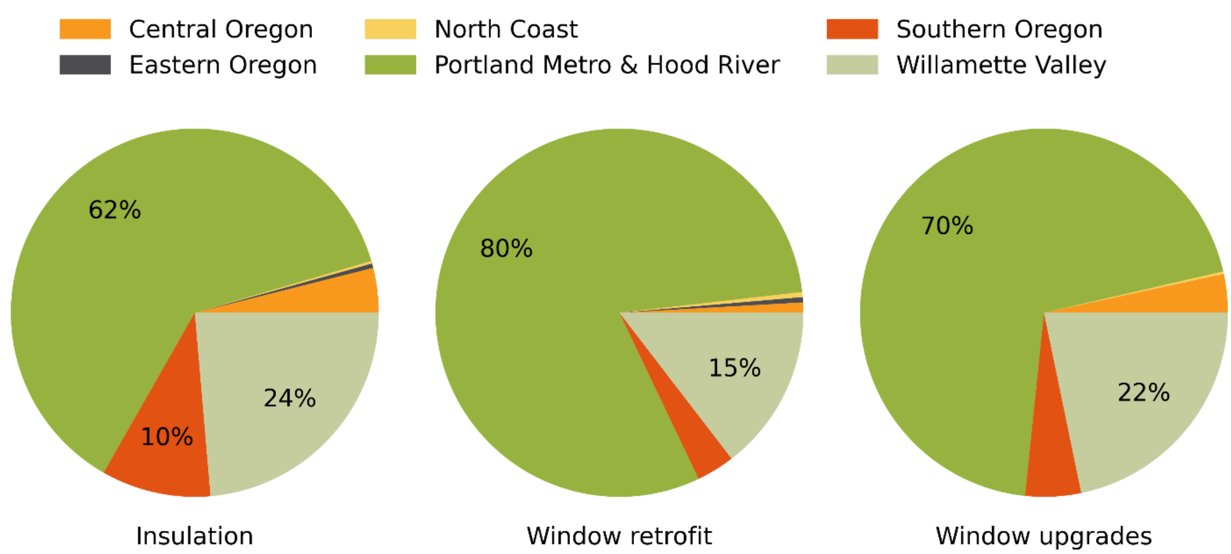
2.2.4 Claimed Savings by Energy Trust Region

[3] Window retrofits in gas heated buildings where not observed in the tracking data during this time period. This measure is only applicable in large multifamily buildings, for which gas heat is quite rare.

[4] The window claimed kWh savings per square foot are relative to market baselines for window upgrades and relative to existing conditions for window retrofits contributing to higher claimed savings per square foot for retrofit windows.

Figure 2-1 summarizes the program tracking savings by region within the Energy Trust service territory. A vast majority of claimed savings in the program are found in the Portland Metro & Hood River (between 62 and 80 percent) and the Willamette Valley regions (between 15 and 24 percent).

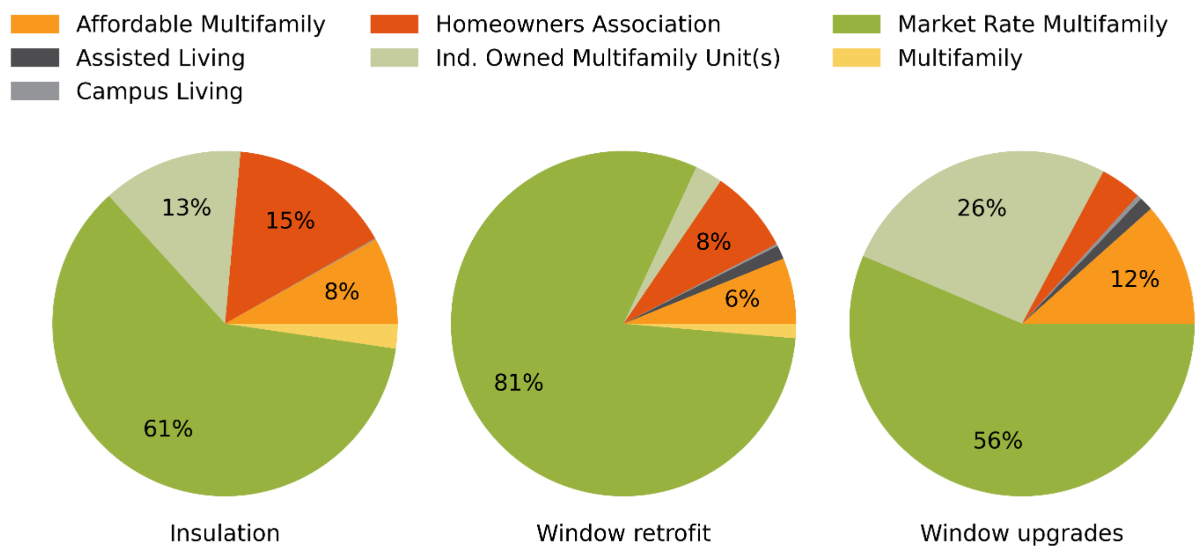
FIGURE 2-1: MULTIFAMILY WEATHERIZATION PERCENTAGE OF CLAIMED KWH SAVINGS BY REGION



2.2.5 Claimed Savings by Energy Trust Market Type

Figure 2-2 summarizes the program tracking claimed savings by Energy Trust Market Type. A majority of claimed savings from the program are in the market rate multifamily market segment (between 56 and 81 percent). The second highest proportion of savings varies between measure type, with window upgrades having a quarter of savings in individually owned multifamily units, and window retrofits and insulation having eight and 15 percent in homeowners' associations, respectively. Affordable multifamily is the third highest proportion of savings for both window measures, with 12 percent of the window upgrade market and six percent of the window retrofit market.

FIGURE 2-2: MULTIFAMILY WEATHERIZATION PERCENTAGE OF CLAIMED KWH SAVINGS BY MARKET TYPE



2.3 EVALUATION OBJECTIVES

The primary objective of this evaluation is to assess the aggregate program electric and natural gas savings from incentivized multifamily windows and insulation measure installations by various domains of interest. Savings are categorized by measure type (window retrofits, window upgrades, and insulation), heating fuel (electric and gas), and building size (small and large). In addition to the overall estimates listed above, other domains of interest include year installed (2016 to 2021), unit ownership (own vs. rent), heating zone (1 vs. 2), and geographic region. Energy savings and realization rates for these domains are estimated when sample sizes and available data allow.

Table 2-4 summarizes the Energy Trust's key analysis scenarios that had expected savings. Many of these scenarios had small sample sizes that were not suitable for billing analysis following data cleaning.[5] The reduction in sample size due to data quality control is described later.

5 The electric savings for windows upgrades and insulation installed in gas heated units was not analyzed due to the small size of savings due to load.

TABLE 2-4: KEY IMPACT ANALYSIS SCENARIOS FOR MULTIFAMILY SHELL MEASURES—EXPECTED FUEL SAVINGS BY MEASURES TYPE, HEATING FUEL, AND BUILDING SIZE

Measure Type	Heating Fuel	Building Size	Project N	Fuel Analyzed	Expected Savings / Measure SqFt (Avg. Claimed)	Evaluable?
Window retrofits	Electricity	Large	496	Electricity	6.87064	Yes
Window upgrades	Electricity	Small	1,194	Electricity	4.07330	Yes, with adj. factors
	Gas	Small	759	Electricity	0.02937	Yes, with adj. factors
				Gas	0.30054	
Ceiling insulation	Electricity	Large	26	Electricity	0.45949	No
		Small	275	Electricity	0.78017	Yes, when combined with wall and floor insulation measures
	Gas	Large	4	Electricity	0.04932	No
				Gas	0.03304	
		Small	138	Electricity	0.10536	Yes, when combined with wall and floor insulation measures
				Gas	0.05504	
Wall insulation	Electricity	Small	53	Electricity	1.11556	Yes, when combined with ceiling and floor insulation measures
	Gas	Small	51	Electricity	0.02266	
				Gas	0.04283	
Floor insulation	Electricity	Large	21	Electricity	0.42988	No
		Small	148	Electricity	0.67843	Yes, when combined with wall and ceiling insulation measures
	Gas	Large	1	Electricity	0.00143	No
				Gas	0.03000	
		Small	58	Electricity	0.00083	Yes, when combined with wall and ceiling insulation measures
				Gas	0.03525	

Note: Project N's represent the number of completed projects with a particular measure in a given scenario, not the number of multifamily sites with a specific fuel savings or with valid UCI data for that fuel. Valid N's available for analysis may be substantially lower. In addition, some projects included multiple measure types.

VERDANT

3 DATA AND METHODS

3.1 DATA SOURCES

This section describes the data sources that were used for the evaluation. There were other data sources examined but were ultimately not used as part of the analysis.

3.1.1 Program Tracking Data

Energy Trust provided Verdant with a Program Tracking (PT) dataset that included participants from program years 2016 to 2023 and included information about each project's installed measures, incentives, energy savings, project address, site id, year constructed, square footage, space heating fuel and system, water heating fuel and system, number of stories, number of dwelling units, heating and cooling zone, weather station, and the installation date. These data include site information that described how multifamily sites are configured, including identifiers for properties with multiple unit buildings (campuses or structures) and individual dwelling units that are attached or located within a building. The PT data also included details about Energy Trust-funded efficiency measures installed prior to, and since, the weatherization measures were installed.

Projects in the tracking database can represent measures installed at a single unit in a multifamily complex and multiple units at a building, complex, or campus.

3.1.2 Customer Relationship Management Data

Energy Trust provided Verdant with a Customer Relationship Management (CRM) dataset. The CRM data contains information about site units in the multifamily complex that can help tie a project in the program tracking database to customer utility bills; including the property's address, utility, site id, premise id, meter number, and a net metering flag.

3.1.3 Weather Data

Energy Trust provided Verdant with daily NOAA weather data and TMYx[6] data for all the weather stations represented in the population of program units from late 2014 to early 2023. The daily NOAA data included actual daily minimum, maximum, and average temperature by weather station or heating zone.

---

[6] TMYx data are typical meteorological year data derived from hourly weather data through 2021 in the ISD (US NOAA's Integrated Surface Database) using the TMY/ISO 15927-4:2005 methodologies.

3.1.4 Utility Customer Information (Gas and Electric Bills)

The PT and CRM data were used to request Utility Customer Information (UCI) data, comprised of monthly electricity and gas consumption, for participants who had installed weatherization measures in multifamily buildings from 2016 into 2023. These data included monthly electric and gas usage from late 2014 through early 2023.

3.1.5 Heating Zones

Energy Trust provided Verdant with heating zones, which are geographic areas defined by the Regional Technical Forum,[7] based on the number of heating degree-days during a typical winter. Heating zone 1 represents areas of the state with relatively mild winters, such as the valleys and coastline of Western Oregon. Heating zones 2 and 3 represent areas of the state with cold winters, like the mountains and Central and Eastern Oregon.

3.1.6 Tables from the Statement of Work (SOW)

Using tables 1 through 4 provided in the project SOW, Verdant used the per square footage deemed savings values and measure information to merge building size (small or large) onto the tracking database. A “small” building is classified as “2 to 4 dwelling units or a side-by-side configuration” and a “large” building is classified as “5 or more dwelling units in a stacked configuration”.

3.1.7 Files for Documentation Review

In addition to program tracking data, Energy Trust provided Verdant with several types of documents related to the program’s implementation. These documents included the program implementation manuals, measure approval documents (MADs), and measure change documents. Verdant reviewed the MAD documents to better understand the development of the ex ante savings values for the various measures. The MAD documentation, previous evaluation reports, and baseline data from the Residential Building Stock Assessment were used to develop an adjustment factor to update the ex ante savings for window upgrades to be consistent with an existing conditions baseline.

3.2 ANALYSIS METHODS

To assess the electric and natural gas savings from the weatherization measures incentivized and installed in Multifamily buildings, the following steps were implemented:

- Data cleaning and quality control including aligning program tracking, billing, and weather data

7 The RTF is a technical advisory committee to the Northwest Power and Conservation Council established in 1999 to develop standards to verify and evaluate energy efficiency savings. https://rtf.nwcouncil.org/

- Determining a unit of analysis
- Developing a matched control group
- Implementing statistical methods to estimate energy savings

3.2.1 Data Cleaning and Quality Control

Data review and cleaning are necessary to identify and resolve data quality issues in preparation for the impact analysis. Types of data cleaning included: flagging any anomalies, removing duplicate records, and properly coding and handling missing values, missing observations, and errors, among other activities. To ensure a comprehensive and accurate dataset of participating multifamily buildings, various datasets were combined. This involved: correctly determining building configurations, connecting the appropriate meters to buildings and dwelling units, and ensuring that the project and site data are aligned.

Combining Program Tracking and Customer Data

Conceptually, the objective of this task is straightforward. Each participant in the program tracking data needs to be cleanly associated with its utility monthly billing records so that it can be used in the subsequent modeling of energy savings. In practice, however, there were a variety of complications at multiple stages of this process that led to ambiguity and uncertainty, rendering a large share of the data provided unsuitable for billing analysis.

The first stage of this task was the merging of the Program Tracking (PT) data with Energy Trust's Customer Relationship Management (CRM) data, which provides the key to mapping to the utility billing data. Both PT and CRM data share a "site_id" column for merging, but the linkage was incomplete, so a substantial portion of records needed to be linked by other means, which left only the address information in the two sources. Because these fields were formatted differently – and because multifamily addresses are notoriously more varied – there was substantial effort made to clean and standardize the various address fields. The results of the merging of the PT and CRM data are presented in Table 3-1, which show that the merge could not find candidate matches for only eight site IDs from the PT data.

TABLE 3-1: MERGING TRACKING DATA TO CUSTOMER DATA

Stage	Site IDs	Percent of Total
Unique values in project tracking data	2,655	100%
- Merged to the customer population by site_id	1,949	73.4%
- Merged to the customer population by cleaned street address and zip code	690	26.0%
- Merged to the customer population by first 15 characters in cleaned street address and zip code	5	0.2%
- Merged to the customer population by first 15 characters in cleaned street address (zip did not match, but from a manual look up appeared to be the same address with bad zip code)	3	0.1%
Unable to merge	8	0.3%

The overall success of this merge, however, belies that there was still a substantial amount of uncertainty in the reliability of the matches. For example, in the PT data, there were cases where a project site had an address range (e.g. 111-119 1st Street). If we were able to merge on any of the address numbers in the range, we counted this as a successful merge, but the lack of unambiguous street and unit numbers in this process introduced substantial uncertainty about how many of the individual units or buildings at a given address were associated with the measures recorded in the PT data. Moreover, some of the tracking records with address ranges had a relatively small amount of square footage installed, implying that the listed addresses may include units or buildings that were not actually treated through the program. This issue was not limited to only tracking records with an address range. For example, there were other tracking records (mainly the "Campus" or "Structure" site types) that had over 100 units and under 100 square feet installed, implying that each unit had, on average, less than one square feet of windows or insulation installed through the program or there were some units where no windows or insulation installed through the program. The most likely case with this example is that the measure was installed in a way that impacted only a subset of the 100 units.

It is also important to note that we were unable to group units into a specific building with the tracking data that was collected for this program. For example, if a Campus had eight buildings, we did not have a way to map each unit to a specific building to be able to analyze each structure separately. Instead, the eight buildings had to be combined to be usable in the analysis.

Assessing the Customer Billing Data

The next stage of this task was the extraction of the billing records. The billing usage was pulled from the Utility Customer Information (UCI) database and then linked back to the PT data that was merged to the CRM records. There is a premise ID available in both the CRM and UCI data, but not every premise ID identified in the CRM data had a match in the UCI. Table 3-2 presents the number of site IDs where we received billing data by analysis segment.

TABLE 3-2: MERGING CUSTOMER DATA TO UTILITY BILLS

Segment	Count in the Program Tracking Data	Count with Billing Data	Percentage with Billing Data
Site IDs with electric window savings	1,424	1,027	72%
Site IDs with gas window savings	563	433	77%
Site IDs with electric insulation savings	476	299	63%
Site IDs with gas insulation savings	176	119	68%

Note – site id is a variable from the PT data. Site id potentially has a one-to-many relationship with the utility billing data.

Note that in addition to the substantial proportion of site IDs from the program tracking data that did not have any billing data records, the identification of billing records does not necessarily mean that the available data was sufficient for use in the analysis. Data cleaning issues are discussed below.

Cleaning the Customer Billing Data

The electric and gas monthly billing records were cleaned, and attention was paid to ensuring that mapping of PT data to bills was treated correctly. Steps in cleaning the billing data included removal of all records where both the kWh and therms were either missing or zero and removal of duplicates. Additionally, there were a number of records that contained identical bill dates and consumption values for a premise that were attached to two different accounts when there was a changeover in the account number and account holder name. This indicated duplicate bills that were associated with two customers during a move-in/out at the premise. In these situations, the bills were removed for the new account (the move-in account). The cleaning also dealt with anomalies in the data, such as excessively long billing intervals.

Calendarizing the Customer Billing Data

Monthly billing records capture the consumption between two dates with an interval that is typically around, but not necessarily, 30 days. Rarely are they associated with a distinct calendar month (e.g., June 1 through June 30), and if so, it is likely by chance. Because a meter read date of, say, July 6th is more representative of the consumption for June, not July, Verdant applied a process called calendarization to the billing data so that consumption is transformed to represent calendar months. The calendarization process essentially allocates the energy consumption in a monthly billing interval to the different months represented in the bill. For example, a bill from June 7th to July 6th would have its electric or gas consumption allocated proportionally to each of those months. This allocation is sometimes done based on the number of days in each month, but Verdant has found that calendarization of weather sensitive loads (such as Multifamily housing) based on heating and cooling degree days (HDD and CDD) provides a more accurate calendarization of bills than allocating energy consumption by the proportional count of days.[8] As a result, billing periods with start dates in winter months (November through March) had their usage allocated to each month based on the relative share of total HDD in each month covered by the billing cycle. Summer months (May through September) usage was allocated based on the relative share of total CDD. For billing cycles that started in April and October (shoulder months) the bills were allocated based on the proportional share of days rather than degree days.

Comparing the Number of Units in the Program Tracking Data to the Customer Billing Data

Aligning the records from the program tracking data to the utility billing data at the unit level led to several complications for the evaluation. First, the number of utility bills didn't always align with the number of units in the tracking data. It was not uncommon to see the number of units in the program tracking data exceed the number of utility billing premise records (implying that we may not have received all of the desired billing records). In other cases, the number of units in the program tracking data was less than the number of utility billing records (potentially implying that we received bills for buildings within a complex that were not treated or bills for common areas or central systems).

We manually compared the tracking data to the billing data to assess the alignment of the number of multifamily units in each data source. At this stage we identified that the records with a site type of "Campus" or "Structure" had more issues arise with the matched program tracking and utility data than records with the site type of "Unit." Among the issues encountered far more frequently with the "Campus" and "Structure" records were the following:

- We did not receive data for each unit for the analysis period
- Some customers moved in or out during the analysis period[9]
- Some units were vacant during the analysis period
- The tracking data did not reflect the same number of units as the bills
  - Multiple addresses in the PT data but fewer bills
  - Range of addresses in the PT data matched with too many bills

Given these data concerns, data for participants from "Campus" and "Structure" records required additional data cleaning and review to develop a measure of ex ante savings and per unit average usage. Because we had some "Campus" and "Structure" records with installations occurring across multiple units and some of the units were missing all bills or their bills were dropped in the cleaning process, we applied

---

8 For the calendarization of customer usage, we used 60 degrees as the change point for HDD and CDD for all premises. This approach minimized the number of months with no HDD or CDD, limiting the proportional allocation of usage by number of days.

9 Installations were often occurring at the point of customer turnover. The change in energy usage at these sites represented changes in customers and the weatherization measure, potentially biasing results if these records were included in the analysis.

an adjustment factor[10] to the aggregate savings to account for the missing bill(s). After applying that adjustment factor, we calculated the pre and post average annual energy consumption per unit by dividing the total usage for the Campus or Structure by the number of units in the PT data. If this appeared unreasonably high or low, we divided the total usage by the number of units with bills. If this still didn't appear reasonable, the site was dropped from the analysis. It is important to note that it is not necessarily the case that all units within the property were treated in the same way, even though we have treated them as though they were all treated the same way. Because the PT data records the total quantity of the measure installed but does not stipulate which units received a measure and in what quantity, this was what the data allowed.

3.2.2 Creating an Analysis Unit

We created an "Analysis ID" that was equivalent to the site ID for records with a site type of "Unit" and combined site IDs with similar addresses when site type was equal to "Structure" or "Campus." This results in a single Analysis ID for each installation occurring at a single unit and a single Analysis ID for all apartments within a structure or campus receiving a weatherization measure.

3.2.3 Identify Comparison Group

A comparison group is a key component to this evaluation's analysis approach. The initial evaluation plan intended to use future program participants as the control group, so that a participant in 2016 would be matched with participant from 2018 or later. For later program years, however, the ability to find matches from the dwindling pool of available future participants was resulting in excessive loss of participants. As a result, the approach was modified to use past program participants as the control for the later year participants (e.g. a 2016 participant as the control for 2018).

For the matching routine, each Analysis ID was aligned with a list of candidate Analysis IDs that were matched on heating zone, building size category, and the measure types installed. Eligible candidates were also screened for any additional projects or upgrades that could significantly impact consumption during the analysis period. The eligible candidates were ranked based on the similarity of their annual consumption for the pre-installation period and the top 15 candidates were preserved for use in the later cleaning and modeling stages. An important nuance to note in this is that some participants could appear twice in the analysis if their program participation was so far apart that it allowed for multiple discrete pre- and post-installation periods. These cases of multiple installation periods were few, but they were

[10] The adjustment factor was created by comparing the number of premises pre data cleaning to the number of premises post data cleaning. The adjustment factor is equal to the number of total premises at the property pre data cleaning divided by the number of premises post data cleaning. For example, a property with four premises pre data cleaning and three post would have an adjustment factor equal to 4/3. This factor is multiple by the sum of usage for premises that made it through the data cleaning process to create the estimate of property level usage.

retained in the analysis to maximize the use of the available data. In some cases, no control participant could be identified based on heating zone, building size category, and the measure types and were dropped from the analysis.

Another important item to note is that finding an eligible comparison candidate became increasingly difficult as the number of units within an Analysis ID increased, specifically for “Campuses” and “Structures”. More units within an Analysis ID led to more chances of the data becoming unusable as participants move, units became vacant, or when a full dataset covering all units within the property was not available. As previously mentioned, in many cases we also do not know with certainty which units and buildings were treated at an Analysis ID. Given these issues, we were unable to find eligible comparison candidates for a vast majority of Analysis IDs that had more than 10 units in the program tracking data.

3.2.4 Impact Analysis

The modeling of measure savings was based on difference-in-differences (DiD) approach based on the follow stages:

1. Weather-normalization of pre- and post-installation monthly bills for all participant and control customers.
2. Regression model to remove time and customer effects from the weather-normalized consumption.
3. Standard DiD model estimation.

The second stage of analysis was important because there was no distinct period for measure installation. Instead, participants had measures installed throughout each program year, often at different times, in no particularly discernible pattern. The second stage was included to address the potential bias that can result from this staggered participation. These three stages are discussed in more detail below.

Stage 1: Weather Normalization

The weather normalization step, which represents the first formal stage in the impact estimation modeling, was the development of weather-normalized estimates of consumption for the pre- and post-installation periods for each Analysis ID. This stage used regression to model the calendarized consumption data on various cooling degree-day (CDD) and heating degree-day (HDD) permutations to determine the outdoor temperature points that best predict customer load. For HDD, the thresholds used were 50, 55, 60, and 65. For CDD, which only applied to electricity, the thresholds used were 60, 65, 70, and 75, and none (no CDD term included). For the electric models, the cooling threshold had to be greater than the heating, which was based on the notion that there is a range of temperatures where no heating or cooling is needed in a house.

Selection of the final model for each Analysis ID (either participant or the individual control candidates) and installation period first filtered for models that produced positive and statistically significant parameter estimates for the valid CDD/HDD degree-day variables. For example, a model that had statistically significant CDD and HDD parameters would be selected over others with only one or no significant parameters. The next criterion was to select those that had positive parameter estimates, whether significant or not. From these candidate models, the model which produced the highest adjusted R-squared value was selected as the final.

After selecting the final CDD and HDD thresholds for each Analysis ID, period (pre/post) and role (participant/control), we applied the parameter estimates to the appropriate TMY data to generate a monthly series of normalized consumption. As a means of quality control and validation, we saved both the actual and weather-normalized consumption values and degree-days. It is important to check for cases where normalized and actual values deviate by extreme amounts, which might indicate an error in the weather data used in either the modeling or normalization procedures.

There was attrition at this stage of the analysis as well, as not every model produced reasonable results, including models with very poor fit, negative degree-day parameters, or weather normalized estimates for pre and post periods that were too different to be believable. While a participant with such cases had to be eliminated, this was one of the reasons for retaining up to 15 control candidates for each participant. If the top control candidate(s) had bad model results, the next best match was selected as the alternate and final control group customer.

Stage 2 and Stage 3

As explained above, the staggered nature of program participation over many years can lead to potential biases, primarily from effect heterogeneity and/or the violation of assumptions of parallel trends. The former is essentially when the effect of treatment can vary over time, which could occur for any number of theoretical reasons, such as earlier participants being larger consumers. The latter is the assumption that treatment and control are both trending in the same direction and at the same rate in the dependent variable. To remedy the potential bias, the analysis employed the “did2s” package in R, which was developed specifically to address these issues. Using the weather-normalized data from stage 1, this package first executes a model to remove the time and site effects due to staggered implementation in the first stage, followed by standard DiD estimation in the second stage. The combination of the weather normalization with these later two stages is referred to as three-stage DiD.

3.2.5 Summary of Exclusion Criteria

There have been references throughout Section 3.2 to data attrition in each step of the data preparation. This section presents a complete view of the data attrition that occurred at each stage of data preparation and analysis.

Table 3-3 presents a high-level attrition table summarizing the number of Analysis IDs that were left after each major data preparation step. In the end we were left with around 30 percent of electric heat Analysis IDs and 35 percent of gas heat Analysis IDs going into our models. We initially requested data for 1,774 Analysis IDs with electric heating and 693 Analysis IDs with gas heating. We received at least some billing data back for 1,255 Analysis IDs with electric heating and 524 Analysis IDs with gas heating, resulting in a drop of over a quarter of the Analysis IDs. Control candidates were assigned to each participant, with participants being dropped if they could not be assigned any suitable controls, resulting in a drop of around 16 percent of the Analysis IDs. The number of units were compared between the PT data and the billing data received. We assessed that the sites that had more than 10 units in the tracking data were nearly all unusable because of the mismatch in number of units for which we had bills during the analysis period or the inability to find a good control candidate; this resulted in a drop of around seven percent of the Analysis IDs. Another 14 percent of the data was removed in the billing data normalization step of the DiD model (this step checks for 12 months of pre- and post- bills). The resulting normalized data were reviewed to identify any issues that could affect the billing analysis, and another five percent of Analysis IDs were removed from the analysis in that step for reasons including low or abnormal usage. Note that some Analysis IDs had multiple projects that made it through to the DiD model, so the sample sizes reported in the models will be slightly higher than the numbers reported here.

TABLE 3-3: ATTRITION BY ANALYSIS ID

Data Processing Step	Electric Heat Analysis IDs	Percent of Electric Heat Analysis IDs Removed	Gas Heat Analysis IDs	Percent of Gas Heat Analysis IDs Removed
Requested Billing Data	1,774		693	
Received Billing Data	1,255	-29%	524	-24%
Control Group Matching	1,015	-14%	381	-21%
Mismatched Number of Units Between PT and Bills	867	-8%	367	-2%
Billing Data Normalization	603	-15%	291	-11%
Final Modeling Filters**	529	-4%	246	-6%

**The final modeling filters included items such as: insufficient billing data, bill outliers, keeping only electric bills for Analysis IDs reporting kWh savings, keeping only gas bills for Analysis IDs reporting therm savings, removing Analysis IDs with 2022 installs, and removing Analysis IDs occurring in a single unit that mapped to multiple billing data premises.

In the end we were left with around 30 percent of electric heat Analysis IDs and 35 percent of gas heat Analysis IDs going into our models. Finally, the Analysis IDs where site type is “Unit” are assigned a weight equal to one and the site types of “Campus” and “Structure” get a weight equal to the number of units in the tracking database.

4 FINDINGS

4.1 DOCUMENT REVIEW

Verdant reviewed the program documentation provided by Energy Trust, including program implementation manuals, measure change documents, and measure approval documents (MADs). Overall, these documents provided valuable background on the program, but the MADs were the only source that were used directly in producing the results presented in this report.

The MADs, which gave key assumptions underlying the measure savings assumptions, were necessary to develop adjustments to the estimated savings for small multifamily window upgrade measures. The MAD's ex ante savings for windows installed in small buildings were based on participants installing higher efficiency windows than a market baseline U-factor. Savings estimated using a billing analysis approach, however, are estimated relative to existing conditions. The information in the MAD, along with data from the 2022 Residential Building Stock Assessment (RBSA)[11] were used to develop the necessary adjustment factors. The RBSA listed the average U-factor for multifamily housing as 0.45. Given that the windows replaced by the Multifamily Weatherization program likely represent older windows than average, we chose to use a U-factor of 0.50 to adjust the market baseline savings to existing conditions. Using the adjusted U-factor it was possible to calculate existing conditions ex ante savings values for window upgrades and calculate an evaluation realization rate for this measure. Table 4-1 presents the deemed market baseline and the adjusted existing baseline ex ante savings values.

11 Evergreen Economics (2024). *2022 Residential Building Stock Assessment Findings Report*. Northwest Energy Efficiency Alliance.

TABLE 4-1: SMALL MULTIFAMILY WINDOW UPGRADE EX ANTE SAVINGS ADJUSTED TO EXISTING BASELINES

Small Multifamily Window Upgrade	Year	Deemed kWh/sqft	Market Baseline U-Factor*	Adjusted kWh/sqft with U = 0.5	Deemed Therm/sqft	Adjusted Therm/sqft with U = 0.5
- **U 0.30 to 0.28**	2014-2020	2.86	0.334	15.64	0.20	1.08
- **U 0.30 to 0.28**	2020-2022	1.84	0.317	15.64	0.13	1.08
- **U < 0.28**	2014-2020	6.92	0.334	19.40	0.48	1.34
- **U 0.25 to 0.27**	2020-2022	3.87	0.317	17.67	0.27	1.22
- **U < 0.25**	2020-2022	6.66	0.317	20.45	0.46	1.41

*The Market Baseline U-Factor for years 2014-2020 comes from MAD28.1 and 28.2 Residential High Performance Windows the 2020-2022 values are from MAD 28.3.

4.2 DIFFERENCE IN DIFFERENCE MODEL SAVINGS

Three-stage DiD models were estimated for Analysis IDs[12] installing window upgrades or retrofits and those installing insulation. These groups were further disaggregated into Analysis IDs with only electric savings and those with electric and gas savings. For Analysis IDs with gas heating, the gas savings were evaluated in the DiD models. The electric savings for these Analysis IDs were not evaluated, as the impacts for these were generally very small and would not be identifiable in an electric model. The DiD models include data for measures installed in Analysis IDs with up to nine units and were implemented on a per unit basis to reduce the heterogeneity of the energy usage variable (monthly kWh or therms) for Analysis IDs with multiple units that were treated. The dependent variable is the average unit level monthly usage, and the model implementation is weighted so each observation represents the number of units receiving treatment.[13] The weighted per unit savings estimated from the model are then compared to the weighted per unit ex ante claimed savings along with the square footage of the installed measures to calculate realization rates. As highlighted in the document review, we developed adjustment factors for the ex ante savings for small multifamily window upgrades to compare estimated savings to both existing conditions (adjusted ex ante) and to market baseline (unadjusted ex ante).

Table 4-2 presents the DiD model estimated savings and resulting observed and adjusted realization rates. The observed realization rates are relative to the program's ex ante savings values. The ex ante savings values for window retrofits and insulation are relative to existing conditions while window upgrades are relative to a market baseline. For window upgrade measures we calculated adjusted realization rates by adjusting the ex ante savings values to be relative to existing conditions. These realization rates are estimated using data from all program years covered by this evaluation (2016 to 2021). For window

12 See section 3.2.2 for a description of this analysis unit.

13 Parameter estimates and other statistics for the third stage of the DiD model are presented in Appendix C. In this appendix, we also describe how the monthly coefficient estimates, and their standard errors, were aggregated to develop the savings estimates presented in Table 4-2.

upgrades, we estimate adjusted realization rates of 34 percent and 10 percent for windows installed at Analysis IDs with electric and gas savings, respectively. For window retrofits, we estimate a realization rate of 65 percent for electric heated Analysis IDs. For insulation measures, we estimate a 67 percent realization rate for electric heated and 90 percent realization rate for gas heated Analysis IDs. For window retrofits and insulation within gas or electric heated multifamily Analysis IDs, while the estimated realization rates are less than 100 percent, their confidence intervals do include a 100 percent realization rate, implying that the ex ante savings values for these measures are not entirely inconsistent with the study's measured values.

The estimated electric energy savings per square foot for window upgrades and window retrofits are 5.61 kWh and 4.87 kWh, respectively. These savings are estimated relative to an existing conditions baseline. The similarity in the savings estimates for upgrades and retrofits implies that the two measures are achieving similar reductions in HVAC energy usage. The similarity in upgrade and retrofit window savings is consistent with the a priori assumption that new windows with similar U-Factors should achieve similar energy savings. The lower adjusted realization rate for window upgrades is from de-rating their savings consistent with the program's implementation style. For window upgrades, the program intervenes, encouraging a higher efficiency window, after the customer has identified that they are going to replace the window. The program's window upgrade savings are therefore assessed relative to a market baseline counterfactual, leading to a lower adjusted realization rate.

TABLE 4-2: ESTIMATED SAVINGS AND REALIZATION RATES SEGMENTED BY MEASURE, HEATING TYPE, AND SIZE

Energy	Measure	Building Size	Model n***	Estimated Energy Savings per Sqft Installed**	# of Months Statistically Significant at 10%	Realization Rate Observed	Realization Rate Adjusted
Electric Heat							
kWh	Window Upgrade	Small	320	5.61 ± 2.08	7	133% ± 49%	34% ± 12%
kWh	Window Retrofit	Large	100	4.87 ± 3.14	5	65% ± 42%	
kWh	Insulation	Small	140	0.45 ± 0.42	5	67% ± 63%	
Gas Heat							
therm	Window Upgrade	Small	229	0.12 ± 0.09	6	40% ± 30%	10% ± 8%
therm	Insulation	Small	35	0.04 ± 0.02	3	90% ± 47%	

**For this and other columns with the "±" following the value, the second value indicates the 90 percent confidence interval developed using the delta method.

***The "Model n" is the total number of install periods for each Analysis ID that went into the model. See section 3.2.2 for a description of Analysis ID.

The savings estimates presented above were developed using data for measures installed in Analysis IDs with up to nine units. Measures installed in Analysis IDs with 10 or more units were not included due to the significant data issues described in Section 3.2 above. Due to substantial data attrition, the results presented above may not be representative when applied to the entire participant population and must be applied with caution.[14]

Approximately 90 percent of the window upgrade Analysis IDs, and approximately 60 percent of the ex ante savings, are associated with complexes with nine or fewer units (see Table A-1 below). This supports the use of the window upgrade model estimates for complexes with nine or fewer units. It is unclear, however, if the findings from these smaller complexes can be applied to window upgrades in complexes with 10 or more units (whose data were excluded from the model). Findings in Table A-3 below show that the average apartment square footage for window upgrade units in building with nine or fewer units was 1,069 while apartments in complexes with 10 or more units were substantially smaller, averaging 660 square feet. The substantially larger average apartment unit square footage of the nine or fewer unit complexes could bias the estimate of savings from smaller apartments upgraded in the 10 plus unit buildings. Therefore, while the modeling of window upgrade savings is based on only a subset of eligible complexes (those with nine or fewer units) and there is significant data attrition within the up to nine unit data, the team feels that the model estimate of savings for window upgrades is sound for units installed in complexes with nine or fewer units.

Ninety three percent of window retrofit savings and 44 percent of Analysis IDs are associated with complexes with 10 or more units. Given the substantial share of window retrofit savings not accounted for in the estimated savings presented in Table 4-2, it is likely that these estimates do not reflect a sound estimate of the savings for this measure and population.

Fifty nine percent of electric insulation savings and 74 percent of gas savings are associated with complexes with nine or fewer units. The electric and gas model insulation savings estimates, however, have either large confidence intervals relative to the savings estimates or are based on a small sample. Caution should be used when applying these savings estimates to the participant population.

Model results by additional domains are included in Appendix B. However, none of the results were statistically significantly different from each other, so these results just serve as a directional indicator of savings within each domain.

[14] See the attrition Table 3-3 for details. The model results presented in Table 4-2 are based on assessment of approximately 30 percent of electric heat Analysis IDs and 35 percent of gas heat Analysis IDs.

Difference-in-Difference Model Caveats

While developing the data for the evaluation of Energy Trust's multifamily weatherization measures, we encountered multiple data quality concerns, many of which were highlighted in Subsection 3.2. Here we further discuss these concerns as they impact the interpretation of the DiD model findings.

First, for Analysis IDs where measures were installed across multiple units at the property, we were unable to consistently determine which specific units within a "Campus" or "Structure" were actually treated and instead had to rely on the assumption that all units were treated equally. Realistically, the installation square footage may have differed across units and buildings. This affected the adjusted savings when there were missing bills. Some records with multiple unit installations were missing bills or some bills were dropped in the cleaning process (due to move-in-move-out, vacancies, or other discrepancies in the data). Because we don't know the square footage installed in each unit, we applied an adjustment factor to the aggregate savings to account for the missing bill(s) that assumes all units were treated with the same square feet of windows or insulation.

Second, it is possible that the premise level energy consumption data for the campus and structure properties include untreated units (as evidence by the small amount of claimed savings and quantity on some of the records that had large number of units in the PT data), which would bias the savings and realization rate results downward.[15]

Third, for several reasons, we were not able to model savings for Analysis IDs (campuses and structures) that had 10 or more units referenced in the PT data. Our review of the PT and UCI found many mismatches in the number of units identified in the two sources as well as uncertainty surrounding the number of units with measures installed. Furthermore, there was a large share of bills that were removed during the data cleaning process and a lack of sufficient control group candidates for these participants. It should be noted that while 10+ unit campuses/structures were the minority of sites receiving MF Weatherization measures (15 percent of records in the PT data), these properties represent the majority of the program's kWh ex ante savings (80 percent of savings in the PT data). The inability to align the PT, CRM, and UCI added a level of error and uncertainty to the data such that the evaluated measure level savings installed in 10+ unit campuses/structures was deemed too unreliable to report.

[15] The realization rates for measures installed in properties where the structure is described as a unit and the number of units is one were higher for all measures other than gas windows. This finding is consistent with the likelihood that the analysis with 2-9 unit structures/campuses includes energy consumption data for units that were not treated.

Finally, note that the PT data included electric savings for measures installed in gas heated units (likely associated with air conditioning and/or furnace fan savings). We were unable to evaluate the claimed electric savings for these measures, due to the small size of the savings.

5 CONCLUSIONS AND RECOMMENDATIONS

The confidence interval for the estimated realization rates include 1.0 or 100 percent for each of the multifamily weatherization measures except window upgrades in electric and gas heated units installed in multifamily buildings from 2016 to 2021. Window upgrades for small buildings have an adjusted realization rate of 0.34 for kWh savings and 0.10 for therms savings (when adjusted for an existing conditions baseline). Window retrofits for large buildings had a realization rate of 0.65 for kWh savings. Insulation installed in small buildings had a realization rate of 0.67 for kWh savings and 0.90 for therms savings. We were unable to estimate savings and realization rates for other measure configurations or subgroups, due to low samples sizes or high attrition.

The estimated kWh realization rates are based on participants that represent a small fraction of the total program kWh savings. Despite outsize efforts to minimize data attrition and the exploration of various alternative modeling approaches (e.g., site-specific NMEC for the larger complexes), we were unable to produce defensible estimates of savings for measures installed in 10+ unit complexes/structures (which were around 15 percent of the records in the PT data, 80 percent of the ex ante kWh savings, and 29 percent of the ex ante therm savings). Multiple data quality concerns impeded modeling. The data issues included incomplete billing data, inability to identify the treated apartment units, the lack of alignment between the PT and UCI data, and the difficulty developing controls for larger buildings. For properties with multiple units treated, the PT data contains information on the property and sparse information on which individual units were treated and individual unit installation quantities. We were able to identify billing information for units at the properties, but we were not able to identify which buildings or units were treated. To remedy these hurdles in future program years we suggest the program implementer collect the following data in the program tracking database:

- A record for each individually treated unit, including address, apartment number, the square footage of insulation or windows installed in the unit (and the square footage of the unit itself), the premise id and meter number associated with the unit (that would directly link to the billing data), an indicator to note if the unit is master-metered, the floor(s) the unit is on, and a common identifier to associate the unit with other units that are in the same building.
- For insulation measures that may affect multiple units, it would be useful to know each specific units the insulation would impact. Again, an identifier to associate all other units that are in the same building would be helpful.

- For complexes and structures that are rentals, we recommend collecting an average vacancy rate from the building owner as part of the application process. This would allow for an adjustment to claimed savings reflect vacancy rates.[16]

If linking bills to each unit in larger buildings is infeasible, we recommend sampling a portion of the participants for a metering study. This sample should stratify the units by energy savings to ensure the largest buildings that account for the largest energy savings will be included. This study could also include surveys of the building managers to better understand vacancy rates.

[16] If the analysis is undertaken on a complex where 10 percent of the apartments are vacant, the usage will be 10 percent lower than expected but the ex ante savings will reflect full occupancy. Accounting for average vacancy in the ex ante savings calculation will align the two sets of data.

APPENDIX A ADDITIONAL MODELS ATTEMPTED

This appendix discusses the additional models that were attempted as part of efforts to incorporate more of the program tracking participants into the analysis but did not end up getting used for the final results.

A.1 NMEC MODEL

Properties with 10 or more units were affected to a greater extent by the issues faced in cleaning, finding a control group, and modeling savings using a Difference-in-Difference model approach. The more units covered under an Analysis ID the more difficult it became to find an eligible control. For electric Analysis IDs of the properties that we were able to find an eligible control the selection was very limited and after cleaning and manual review, they were found to not necessarily have data that would be able to align with the pre and post periods of the treated properties. For these reasons, we concluded that these properties could not be modeled with DiD techniques.

However, because this group of participants contributes a large portion of the kWh savings, especially for the window retrofit measure, we explored alternative methods to model these savings. Because we could not identify a control group for these participants, we attempted to analyze these properties using a traditional normalized metered energy consumption (NMEC) approach, in which savings are estimated based on a simple comparison of pre- and post-installation weather normalized savings for the individual participants. This NMEC approach is grounded in the assumption that the given analysis unit has usage that is predictable and driven in large part by weather, which is surely true of the larger buildings. Because no control groups are used, the NMEC model is not able to account for other events that may cause a substantial change in energy usage. A highly salient example is the global COVID-19 pandemic, which significantly altered how home energy was used during the years 2020 and 2021, as people increased their time at home and energy consumption. Without a control group to account for these types of effects, traditional NMEC can produce misleading results. The evaluation period for this analysis is for multifamily sites with installations between 2016 and 2021, leading to the evaluation not being able to control for this change in energy usage for installations occurring in 2019 through 2021. More importantly, NMEC models rely on model metrics that test for the measure of random error between normalization and actual data, bias, and variation. In calculating site level usage, we run weather normalization models on individual premise level data to allow for varying weather sensitivities and preferences. Statistics were calculated for these weatherization models, specifically the fractional savings uncertainty (FSU) for each premise under an Analysis ID. This is calculated as the level of estimation uncertainty divided by the savings, with lower values meaning a more precise estimate. Overall, these statistics indicated that these sites were not good candidates for NMEC analysis, as the noise associated with their billing records was too large to reliably capture the signal of program impacts. These NMEC-specific issues combined with the data reliability issues led us to determine that these properties were not suitable for either DiD or NMEC estimation.

As stated earlier, the ex ante savings vary considerably for the 1-9 unit compared to the 10+ unit records in the PT data. Table A-1 presents the distribution of claimed savings, square feet installed, and number of Analysis IDs for the two groups by measure and building size. For window upgrades and small building insulation, the majority of participants, square feet installed, claimed kWh savings, and claimed therm savings are in the 1-9 unit group. For window retrofits, while a majority of the Analysis IDs are in the 1-9 unit group (52 percent), a majority of the square footage installed and kWh savings are in the 10+ units group (93 percent for both).

TABLE A-1: RECORDS BY MEASURE AND BUILDING SIZE 1-9 UNIT VS 10+ UNIT

Measure	Building Size	Metric	1-9 Units	10+ Units	NA Units	Total
Window Upgrade	Small	Claimed kWh Savings	57%	40%	3%	971,613
		Claimed therms Savings	63%	32%	4%	39,697
		Sq Ft Installed	59%	38%	3%	369,666
		Number of Analysis IDs	88%	6%	6%	1,432
Window Retrofit	Large	Claimed kWh Savings	6%	93%	1%	5,096,756
		Claimed therms Savings				0
		Sq Ft Installed	6%	93%	1%	741,817
		Number of Analysis IDs	52%	44%	4%	434
Insulation	Small	Claimed kWh Savings	59%	38%	3%	727,029
		Claimed therms Savings	74%	21%	5%	16,908
		Sq Ft Installed	61%	35%	4%	1,250,605
		Number of Analysis IDs	88%	7%	6%	527
Insulation	Large	Claimed kWh Savings	3%	97%	0%	273,449
		Claimed therms Savings	56%	40%	4%	519
		Sq Ft Installed	4%	96%	0%	618,998
		Number of Analysis IDs	33%	64%	2%	42

The numbers in this table were calculated based on PT data prior to merging with billing data.

A major difference between the 1-9 unit group and the 10+ group is the overwhelming prevalence of renters in 10+ unit group, while the 1-9 unit group has a mixture of owners and renters, as shown in Table A-2 below.

TABLE A-2: PERCENT OF RENTERS IN 1-9 UNIT VS 10+ UNIT RECORDS BY MEASURE AND BUILDING SIZE

Measure	Building Size	Metric	1-9 Units % Renters	10+ Units % Renters	Overall % Renters	Total
Window Upgrade	Small	Claimed kWh Savings	56%	97%	72%	971,613
		Claimed therms Savings	36%	100%	56%	39,697
		Sq Ft Installed	49%	98%	67%	369,666
		Number of Analysis IDs	40%	99%	42%	1,432
Window Retrofit	Large	Claimed kWh Savings	57%	100%	98%	5,096,756
		Claimed therms Savings				0
		Sq Ft Installed	59%	100%	98%	741,817
		Number of Analysis IDs	31%	99%	61%	434
Insulation	Small	Claimed kWh Savings	69%	72%	70%	727,029
		Claimed therms Savings	62%	100%	68%	16,908
		Sq Ft Installed	67%	76%	69%	1,250,605
		Number of Analysis IDs	55%	94%	57%	527
Insulation	Large	Claimed kWh Savings	66%	100%	99%	273,449
		Claimed therms Savings	95%	100%	93%	519
		Sq Ft Installed	68%	100%	99%	618,998
		Number of Analysis IDs	36%	100%	76%	42

The numbers in this table were calculated based on PT data prior to merging with billing data.

Table A-3 presents some additional characteristics of the multifamily buildings by the 1-9 unit group and 10+ unit group. The age and claimed savings per square foot installed is fairly similar for the two groupings, but the size (in square feet) of the unit is generally smaller for the larger unit installation records when compared to the single-unit records.

TABLE A-3: BUILDING CHARACTERISTICS BY 1-9 UNIT VS 10 + UNIT RECORDS

Measure	Metric	1-9 Units	10+ Units	NA Units	Total
Window Upgrades	Average Age	1972	1970	1960	1972
	Average Square Feet of Unit	1,069	660		1,202
	Average Number of Units	1.6	34.2		4
	Average Number of Floors	1.7	1.6	1.4	1.7
	Claimed kWh Savings/Sqft	2.56	2.74	2.46	2.63
	Claimed therms Savings/Sqft	0.12	0.09	0.23	0.11
Window Retrofit	Average Age	1972	1971	1950	1971
	Average Square Feet of Unit	936	825		905
	Average Number of Units	2	58.3		27
	Average Number of Floors	1.6	2.3	2.0	2.0
	Claimed kWh Savings/Sqft	6.74	6.90	4.70	6.87
	Claimed therms Savings/Sqft	0.00	0.00	0.00	0
Insulation	Average Age	1963	1967	1949	1963
	Average Square Feet of Unit	1,032	696		1,150
	Average Number of Units	1.8	41.5		6
	Average Number of Floors	1.5	1.9	1.6	1.5
	Claimed kWh Savings/Sqft	0.55	0.53	0.47	0.54
	Claimed therms Savings/Sqft	0.02	0.00	0.02	0.01

The numbers in this table were calculated based on PT data prior to merging with billing data.

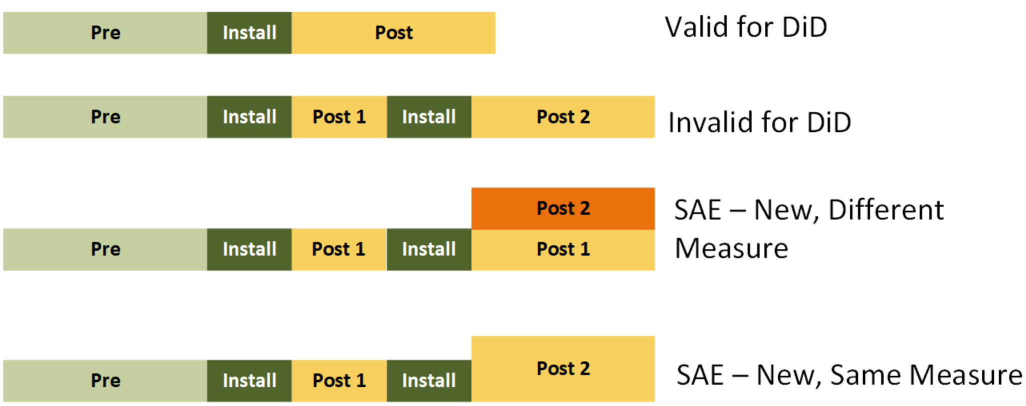
A.2 STATISTICALLY ADJUSTED ENGINEERING MODEL

One concern that arose in the implementation of the three-stage DiD approach was the large number of participants that had to be dropped due to issues with the data. One of the main sources of attrition was the presence of multiple dispersed installation dates for a single unit. In these cases, a participant would have a measure installed and then, for example, six months later, have another measure installed. These additional interventions impeded the ability to identify clear post-installation periods with at least 10 consecutive months.

As a remedy to this problem, we tested using a statistically adjusted engineering (SAE) approach. In an SAE approach, ex ante savings are used directly in a panel data structure as measure-specific independent variables. Instead of interfering with the post period, additional measures installed are simply added to the ex ante savings used in the model. For example, if 10 kWh of windows were installed in January of 2018, in the DiD approach, the installation of 20 kWh of insulation in August of the same year would invalidate the post period. In the SAE approach, the new measure would just become another variable in the model. Similarly, additional windows would just lead to an increase in the expected ex ante savings.

To supplement the description above, **Error! Reference source not found.** provides a visual portrayal of how the SAE approach might mitigate the data attrition due to multiple measure installations. In the top example, there is one installation with an uninterrupted post period, which is the condition needed for the DiD approach. In the next example, an additional installation breaks up the post period into two segments, neither of which is long enough to use individually (note that such cases of multiple 10+ month post periods did occur, and we did use them in the DiD modeling). The bottom two examples show how the SAE allows us to use the data in the model. In the penultimate, the case is where, say, insulation is first installed, followed by windows some months later. For this situation, the insulation ex ante savings are introduced into the model at the first installation period and continue throughout the post (if not clear, they have a value of zero prior to their installation date). After the second installation, window ex ante savings are then introduced into the model in an additional variable. In the final example, the situation is that the same type of measure is installed, in which case their savings are added to the ex ante savings (as represented by the taller segment for "Post 2."

FIGURE A-1: DID VERSUS SAE ATTRITION EXAMPLE



The SAE approach, however, also introduced additional sources of uncertainty to the modeling. First, the annual ex ante savings need to be allocated to monthly values. We did this using heating- and cooling-degree days, which, while defensible and frankly the only feasible method, certainly introduced error into the model.

The next issue with the SAE approach is that the installation dates in the program tracking data are imprecise, so the exact time that savings will be observed is unknown. With the DiD approach, we can apply an ample "dead band" around the installation dates, and then produce weather normalized data for more cleanly identified pre and post periods. With the SAE approach, we cannot exclude that quantity of data from the model without creating a highly imbalanced panel data design (what do time effects mean when a large share of participants does not have data for them?), so this introduced uncertainty from the actual installation timing into the model.

Finally, the SAE models required separate terms for different measures as well as variables for weather. These more complex specifications invariably introduce more noise than found in the DiD weather normalized series, which were also modeled separately by measure type. These are not flaws in the method, but they are differences that can result in marked differences in the modeling results.

In addition to the above sources of uncertainty, the SAE modeling data relied on monthly bills prior to normalization, which required a separate type of data cleaning routine in which it was difficult to maintain consistent exclusion criteria. After going through this cleaning process, the reduction in attrition was not substantial, which was the primary objective of trying this approach.

The SAE model produced results showing program savings, in some cases consistent with those found in the DiD modeling, but not always. This presents the problem of having two sets of results, so given that the increase in sample was not substantial and that this approach deviated from the evaluation plan, we have opted to omit them from this document.

APPENDIX B DIFFERENCE-IN-DIFFERENCE MODEL RESULTS BY DOMAIN

The tables in this section display segmented model results for windows and insulation installations in electric and gas heated units. We only show results where the model N or number of Analysis IDs is at least 30. None of the results were statistically significantly different from the results presented above (and in the All row) or from each other, therefore, these results serve as a directional indicator of savings within each domain.

TABLE B-1: ESTIMATED ELECTRIC WINDOW UPGRADE SAVINGS SEGMENTED BY ADDITIONAL DOMAINS

Energy	Domain	Domain Segment	# Analysis ID / Installation Period	Estimated Energy Savings per Square Feet Installed	Number of Months Statistically Significant at 10%	Realization Rate Observed	Realization Rate Adjusted
kWh	All	All	320	5.61 ± 2.08	7	1.33 ± 0.49	0.34 ± 0.12
kWh	Rent/ Own	Rent	53	6.90 ± 6.55	3	1.56 ± 1.48	0.40 ± 0.38
kWh		Own	264	5.06 ± 1.91	8	1.21 ± 0.46	0.31 ± 0.12
kWh	Region	Portland & HR	262	5.22 ± 2.32	7	1.21 ± 0.54	0.31 ± 0.14
kWh		Willamette Valley	32	0.38 ± 5.31	1	0.10 ± 1.34	0.02 ± 0.35
kWh	Heating Zone	HZ1	315	5.35 ± 2.08	6	1.27 ± 0.49	0.32 ± 0.12
kWh		HZ2	5				
kWh	Year	2016	44	2.94 ± 8.03	0	0.68 ± 1.86	0.18 ± 0.48
kWh		2017	55	7.82 ± 5.35	6	1.60 ± 1.10	0.45 ± 0.31
kWh		2018	50	7.38 ± 4.20	6	1.51 ± 0.86	0.42 ± 0.24
kWh		2019	40	3.55 ± 5.30	0	0.85 ± 1.27	0.21 ± 0.31
kWh		2020	51	7.51 ± 6.63	5	1.81 ± 1.60	0.45 ± 0.40
kWh		2021	80	5.08 ± 3.00	5	1.58 ± 0.94	0.33 ± 0.19

TABLE B-2: ESTIMATED ELECTRIC WINDOW RETROFIT SAVINGS SEGMENTED BY ADDITIONAL DOMAINS

Energy	Domain	Domain Segment	# Analysis ID / Installation Period	Estimated Energy Savings per Square Feet Installed	Number of Months Statistically Significant at 10%	Realization Rate
kWh	All	All	100	4.87 ± 3.14	5	0.65 ± 0.42
kWh	Rent/Own	Rent	11			
kWh		Own	89	6.39 ± 3.22	7	0.87 ± 0.44
kWh	Region	Portland & HR	97	4.49 ± 3.14	5	0.60 ± 0.42
kWh		Willamette Valley	2			
kWh	Heating Zone	HZ1	99	4.67 ± 3.13	5	0.63 ± 0.42
kWh		HZ2	1			
kWh	Year	2016	12			
kWh		2017	35	11.09 ± 6.41	6	1.87 ± 1.08
kWh		2018	16			
kWh		2019	9			
kWh		2020	14			
kWh		2021	14			

TABLE B-3: ESTIMATED ELECTRIC INSULATION SAVINGS SEGMENTED BY ADDITIONAL DOMAINS

Energy	Domain	Domain Segment	# Analysis ID / Installation Period	Estimated Energy Savings per Square Feet Installed	Number of Months Statistically Significant at 10%	Realization Rate
kWh	All	All	140	0.45 ± 0.42	5	0.67 ± 0.63
kWh	Rent/Own	Rent	52	0.21 ± 0.79	7	0.31 ± 1.19
kWh		Own	88	0.65 ± 0.37	9	0.95 ± 0.54
kWh	Region	Portland & HR	103	0.19 ± 0.35	3	0.27 ± 0.50
kWh		Willamette Valley	15			
kWh	Heating Zone	HZ1	140	0.45 ± 0.42	5	0.67 ± 0.63
kWh	Year	2016	17			
kWh		2017	16			
kWh		2018	9			
kWh		2019	19			
kWh		2020	37	1.10 ± 0.70	4	2.03 ± 1.29
kWh		2021	42	0.12 ± 0.55	2	0.23 ± 1.01

TABLE B-4: ESTIMATED GAS WINDOW UPGRADE SAVINGS SEGMENTED BY ADDITIONAL DOMAINS

Energy	Domain	Domain Segment	# Analysis ID / Installation Period	Estimated Energy Savings per Square Feet Installed	Number of Months Statistically Significant at 10%	Realization Rate Observed	Realization Rate Adjusted
therm	All	All	229	0.12 ± 0.09	6	0.40 ± 0.30	0.10 ± 0.08
therm	Rent/ Own	Rent	26				
therm		Own	203	0.08 ± 0.09	2	0.27 ± 0.30	0.07 ± 0.08
therm	Region	Portland & HR	192	0.09 ± 0.11	1	0.28 ± 0.34	0.07 ± 0.09
therm		Willamette Valley	26				
therm	Heating Zone	HZ1	221	0.12 ± 0.10	4	0.38 ± 0.32	0.10 ± 0.08
therm		HZ2	3				
therm	Year	2016	19				
therm		2017	45	0.10 ± 0.14	2	0.25 ± 0.36	0.08 ± 0.11
therm		2018	23				
therm		2019	36	-0.08 ± 0.21	0	-0.24 ± 0.59	-0.07 ± 0.17
therm		2020	50	0.12 ± 0.16	0	0.42 ± 0.59	0.10 ± 0.14
therm		2021	56	0.28 ± 0.17	5	1.16 ± 0.72	0.24 ±0.15

TABLE B-5: ESTIMATED GAS INSULATION SAVINGS SEGMENTED BY ADDITIONAL DOMAINS

Energy	Domain	Domain Segment	# Analysis ID / Installation Period	Estimated Energy Savings per Square Feet Installed	Number of Months Statistically Significant at 10%	Realization Rate
therm	All	All	35	0.04 ± 0.02	3	0.90 ± 0.47
therm	Rent/Own	Rent	7			
therm		Own	28			
therm	Region	Portland & HR	29			
therm		Willamette Valley	4			
therm	Heating Zone	HZ1	35	0.04 ± 0.02	3	0.90 ± 0.47
therm	Year	2016	5			
therm		2017	5			
therm		2018	7			
therm		2019	2			
therm		2020	9			
therm		2021	7			

APPENDIX C REGRESSION RESULTS AND CONFIDENCE INTERVAL CALCULATION

C.1 REGRESSION RESULTS

The regression results from the three-stage DiD estimations are displayed in this section, by measures and fuels. The first stage of the DiD estimation, the weather normalization of consumption, is explained in depth in section 3.2.41. This stage is necessary to remove any fluctuations from normalized consumption brought on by weather. The second and third stages of the DiD are estimated together in a single R package. These final two stages are weighted by the number of units treated at each Analysis ID. For the first part of the final two stages, the model regresses normalized monthly unit-level usage (kWh or therms) on bill date[17] and Analysis ID. This removes any variation in normalized usage caused by time and group effects. Regressing on bill date in this second stage ensures that any outstanding factors caused by time, excluding weather effects that were removed in stage one, are removed. The removal of group and time effects is necessary due to the staggered treatment within the weatherization program, ensuring that the observed changes in consumption are associated with the measures installed, and not some other factor that would bias the result. An extreme example of this would be Covid-19, which would lead to substantial bias in the findings for participants that had measures installed around the onset of the pandemic. For the third stage, the estimated second stage consumption is regressed on a treatment model binary crossed with a month indicator. The treatment model variable is a binary indicator equal to one during the post-installation period for participants and zero for all other periods for participants and all periods for control observations. We present only results from the third stage models within this section as parameter estimates from the first and second stages are not directly relevant to the study question.

Coefficients for treatment crossed with each month are displayed in Table C-1 through Table C-5. Negative estimated coefficients represent program savings while positive coefficients represent increased usage. As expected, months with statistically significant savings tend to be heating months October through April. The number of participants is reported in each table, with each participant having 12 months of pre and 12 months of post data. Additionally, there are an equal number of controls included in each model that also have 12 months of pre and 12 months of post data.

The R-squared and adjusted R-squared for each model are also reported for each analysis. One might note that these values are low for each analysis, but this is associated with the specific approach used to model impacts. The first two stages of the DiD analysis essentially remove weather specific, site specific, and

17 The bills are calendarized to start on the first day of each month, aligning the bill date with the corresponding month and year.

time specific variation from normalized consumption in our models, leaving only the effect of treatment to explain the remaining variation. While R-squared is a measure of the variation explained by the model, these results only reflect the final stage, so one does not see how much variability was accounted for by the first two stages of modeling. Therefore, in DiD analysis it is important to focus more on the coefficients on the treatment interaction terms, as they provide more information on the validity and magnitude of the treatment effects.

TABLE C-1: ELECTRIC WINDOW UPGRADES IN SMALL SITES REGRESSION RESULTS

Dependent Variable	Estimated Normalized Monthly Usage (kWh)	Significance	Standard Error
treatment_model = 1 x month = 1	-97.74	***	21.05
treatment_model = 1 x month = 2	-85.59	***	16.26
treatment_model = 1 x month = 3	-73.50	***	14.96
treatment_model = 1 x month = 4	-41.31	**	12.93
treatment_model = 1 x month = 5	-20.79		13.25
treatment_model = 1 x month = 6	-14.79		15.66
treatment_model = 1 x month = 7	-24.25		19.14
treatment_model = 1 x month = 8	-24.08		19.83
treatment_model = 1 x month = 9	-24.24		15.06
treatment_model = 1 x month = 10	-25.82	.	13.78
treatment_model = 1 x month = 11	-66.20	***	15.41
treatment_model = 1 x month = 12	-98.31	***	19.40
S.E. Type			Custom
Number of Participant Analysis IDs			320
R2			0.0174
Adjusted R2			0.0167

Significance codes: ‘***’ 0.001, ‘**’ 0.01, ‘*’ 0.05, ‘.’ 0.1

TABLE C-2: ELECTRIC WINDOWS RETROFITS IN LARGE SITES REGRESSION RESULTS

Dependent Variable	Estimated Normalized Monthly Usage (kWh)	Significance	Standard Error
treatment_model = 1 x month = 1	-20.78		34.83
treatment_model = 1 x month = 2	-63.72	**	23.00
treatment_model = 1 x month = 3	-59.87	***	17.67
treatment_model = 1 x month = 4	-51.38	***	15.49
treatment_model = 1 x month = 5	-27.63		18.65
treatment_model = 1 x month = 6	-16.26		21.94
treatment_model = 1 x month = 7	-24.36		23.79
treatment_model = 1 x month = 8	-22.26		23.10
treatment_model = 1 x month = 9	-28.51		22.61
treatment_model = 1 x month = 10	-39.07	*	19.62
treatment_model = 1 x month = 11	-39.92	.	20.63
treatment_model = 1 x month = 12	-41.22		29.69
S.E. Type			Custom
Number of Participant Analysis IDs			100
R2			0.01095
Adjusted R2			0.00868

Significance codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1

TABLE C-3: ELECTRIC INSULATION IN SMALL SITES REGRESSION RESULTS

Dependent Variable	Estimated Normalized Monthly Usage (kWh)	Significance	Standard Error
treatment_model = 1 x month = 1	-149.00	***	35.60
treatment_model = 1 x month = 2	-87.56	***	26.13
treatment_model = 1 x month = 3	-46.02	*	22.38
treatment_model = 1 x month = 4	-5.86		19.11
treatment_model = 1 x month = 5	4.66		20.99
treatment_model = 1 x month = 6	8.28		23.37
treatment_model = 1 x month = 7	20.31		29.78
treatment_model = 1 x month = 8	20.35		28.55
treatment_model = 1 x month = 9	13.34		24.32
treatment_model = 1 x month = 10	-6.78		19.49
treatment_model = 1 x month = 11	-77.36	**	24.83
treatment_model = 1 x month = 12	-123.9	***	32.09
S.E. Type			Custom
Number of Participant Analysis IDs			140
R2			0.02871
Adjusted R2			0.02711

Significance codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1

TABLE C-4: GAS WINDOW UPGRADES IN SMALL SITES REGRESSION RESULTS

Dependent Variable	Estimated Normalized Monthly Usage (therms)	Significance	Standard Error
treatment_model = 1 x month = 1	-1.87		1.73
treatment_model = 1 x month = 2	-2.18	.	1.26
treatment_model = 1 x month = 3	-2.26	*	0.98
treatment_model = 1 x month = 4	-2.41	**	0.86
treatment_model = 1 x month = 5	-1.77	.	0.99
treatment_model = 1 x month = 6	-0.67		1.18
treatment_model = 1 x month = 7	-0.20		1.31
treatment_model = 1 x month = 8	-0.49		1.31
treatment_model = 1 x month = 9	-0.67		1.16
treatment_model = 1 x month = 10	-2.06	*	0.96
treatment_model = 1 x month = 11	-2.05	*	1.03
treatment_model = 1 x month = 12	-2.14		1.517
S.E. Type			Custom
Number of Participant Analysis IDs			229
R2			0.00335
Adjusted R2			0.00235

Significance codes: ‘***’ 0.001, ‘**’ 0.01, ‘*’ 0.05, ‘.’ 0.1

TABLE C-5: GAS INSULATION IN SMALL SITES REGRESSION RESULTS

Dependent Variable	Estimated Normalized Monthly Usage (therms)	Significance	Standard Error
treatment_model = 1 x month = 1	-8.67	.	4.61
treatment_model = 1 x month = 2	-4.83		3.26
treatment_model = 1 x month = 3	-3.49		2.18
treatment_model = 1 x month = 4	-1.67		1.86
treatment_model = 1 x month = 5	-0.62		2.04
treatment_model = 1 x month = 6	-1.10		2.62
treatment_model = 1 x month = 7	-2.76		2.60
treatment_model = 1 x month = 8	-2.06		3.60
treatment_model = 1 x month = 9	-1.52		2.67
treatment_model = 1 x month = 10	-0.31		1.97
treatment_model = 1 x month = 11	-6.13	**	2.32
treatment_model = 1 x month = 12	-8.86	*	4.19
S.E. Type			Custom
Number of Participant Analysis IDs			35
R2			0.02483
Adjusted R2			0.0184

Significance codes: ‘***’ 0.001, ‘**’ 0.01, ‘*’ 0.05, ‘.’ 0.1

C.2 INTERPRETTING MONTHLY REGRESSION RESULTS AND CONSTRUCTING CONFIDENCE INTERVALS

The final stage of our approach estimates monthly savings caused by the weatherization measure installations. The annual savings, which are the sum of the individual monthly coefficients, have underlying uncertainty, so it is useful to estimate confidence intervals to help interpret the results. Were the savings generated from a single model coefficient, its standard error could be used directly to calculate the confidence intervals. For this analysis, however, there are coefficients for the individual months, so these need to be aggregated to develop the overall confidence interval for the annual savings. This section describes the method used to generate these confidence intervals.

The standard errors developed to calculate the confidence intervals for these results were based on the delta method. The delta method is a statistical technique used to approximate the standard error of a function of an estimator. It leverages the principles of Taylor series expansion to transform the variance of a simple estimator into the variance of a more complex function of that estimator. Although it is particularly beneficial for nonlinear functions, it also applies seamlessly to linear functions, such as those estimated for this evaluation.

For this evaluation, the function is linear, so the variance of the annual effect is the sum of the variances of the individual monthly effects. The standard error of the annual estimate is the square root of this total variance, and this standard error is then easily constructed as we have an estimate of the standard error of the model.