



Low-Income Energy Affordability Data (LEAD) Tool Methodology

Ookie Ma and Krystal Laymon
U.S. Department of Energy

Megan Day, Ricardo Oliveira, Jon Weers, and Aaron Vimont
National Renewable Energy Laboratory

**NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC**

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Technical Report
NREL/TP-6A20-74249
July 2019



Low-Income Energy Affordability Data (LEAD) Tool Methodology

Ookie Ma and Krystal Laymon
U.S. Department of Energy

Megan Day, Ricardo Oliveira, Jon Weers, and Aaron Vimont
National Renewable Energy Laboratory

Suggested Citation

Ma, Ookie, Krystal Laymon, Megan Day, Ricardo Oliveira, Jon Weers, and Aaron Vimont. 2019. *Low-Income Energy Affordability Data (LEAD) Tool Methodology*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-6A20-74249.
<https://www.nrel.gov/docs/fy19osti/74249.pdf>.

**NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC**

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Technical Report
NREL/TP-6A20-74249
July 2019

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Wind Energy Technologies Office. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

U.S. Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free via www.osti.gov.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

NREL prints on paper that contains recycled content.

Acknowledgments

The authors thank the following people for their contributions: Brian Bush (NREL) for creating the initial data methodology and Kerry O’Neill and Isabelle Hazlewood (Connecticut Green Bank) and Christina Bowman (Fellow at the U.S. Department of Energy) for their guidance during the early stages of tool development. We also gratefully acknowledge funding for this work from the Office of Weatherization and Intergovernmental Programs and the Office of Strategic Programs of the U.S. Department of Energy’s Office of Energy Efficiency and Renewable Energy under contract number DE-AC36-08GO28308.

U.S. Department of Energy would also like to acknowledge the feedback from stakeholders during the development of the LEAD Tool: American Council for an Energy-Efficient Economy, City of Philadelphia, Connecticut Green Bank, Inclusive Prosperity Capital Inc., National Grid Southeast Efficient Energy Alliance, State of Kentucky, and State of Minnesota.

Definitions

Census blocks are the smallest geographic areas used by the U.S. Census Bureau and are bounded by visible features, such as streets, roads, streams, and railroad tracks. Census blocks nest within all other census geographic entities including cities, townships, and counties.

Census tracts are small, statistical subdivisions of a county or a county equivalent. They generally have a population size between 1,200 and 8,000 people, with an average size of 4,000 people.¹

Minor civil divisions are the primary governmental and/or administrative divisions of a county and include townships and towns. In many states, minor civil divisions can serve as general-purpose governments.

¹ U.S. Census Bureau. n.d. *Census Tracts*. <https://www2.census.gov/geo/pdfs/education/CensusTracts.pdf>.

Table of Contents

1	Introduction	1
2	Example Tool Outputs	3
3	Methodology	7
3.1	Household Energy Estimates.....	7
4	Calibration	10
4.1	Energy intensity weighting.....	10
5	Validation of Estimates to Measured Values	12
5.1	Validation of Modeled Data.....	12
6	Discussion	15
	Appendix A. Derivation of Calibration Equations	17

1 Introduction

State and local governments, utilities, non-profits, weatherization providers, and other stakeholders often lack easy access to data to help inform their decision making and have incomplete knowledge of the areas they serve. The Low-income Energy Affordability Data (LEAD) Tool, developed by U.S. Department of Energy and National Renewable Energy Laboratory, helps stakeholders make data-driven decisions on energy goal setting and program planning by providing them information on low-income household populations and associated energy use characteristics.

The LEAD Tool is a web-accessible², interactive platform that allows users to build their own national, state, city, or county profiles with estimated, locally specific low-income household energy characteristics. Users are able to visualize the profiles as well as compare differences among profiles. In addition, users will be able to save their profiles and download visualizations and data.

The LEAD Tool provides interactive maps, charts, and data for housing unit counts; average monthly housing electricity, gas, and other fuel expenditures; and average energy burden³ tabulated by the following categories.

- Geographic level
 - National
 - 50 States plus D.C. and Puerto Rico
 - County
 - City
 - Census tract
- Household income level
 - Area Median Income (AMI): 0-30%, 30-60%, 60-80%, 80-100%, 100+%
 - Federal Poverty Level (FPL): 0%- 100%, 100%-150%, 150%-200%, 200%-400%, 400% +
- Housing unit type
 - Tenure: home owners versus renters
 - Building year of first construction
 - Number of units in the building
 - Housing unit primary heating fuel type

The LEAD Tool can be used by stakeholders to fill knowledge gaps, conduct strategic energy planning, inform the development of low-income energy programs and goals, improve energy program targeting, and increase public awareness of low-income household issues. However, this tool is not meant to be used as a substitute for evaluating program or policy effectiveness or for

² Available online at <https://www.energy.gov/eere/slsc/maps/lead-tool>

³ Energy burden is defined as the average annual housing energy costs divided by the average annual household income.

tracking year-to-year changes. Since LEAD relies upon survey data and estimation techniques, associated uncertainties and margins of errors render a time series representation of the data based on different LEAD Tool vintages unreliable for analysis purposes.

2 Example Tool Outputs

The web-based LEAD Tool enables user exploration of residential household energy profiles. For illustration purposes, a short summary of some of the tool features are provided in this section.

The tool provides three principle metrics: energy burden, annual average housing energy costs, and housing counts. Energy burden is defined as the average annual housing energy costs divided by the average annual household income. Monthly housing energy costs are based on household monthly expenditures for electricity, gas (utility and bottled), and other fuels (including fuel oil, wood, etc.). Housing counts are the number of occupied housing units. As such, the tool assumes that the terms housing units and households are interchangeable. The user may examine these metrics for various housing unit cohorts. Cohorts may be based on geography. Available geographies include national, state (see Figure 1), county (see Figure 4), city, and census tract levels. Cohorts may also be based on housing unit characteristics, which will be discussed in more detail below.

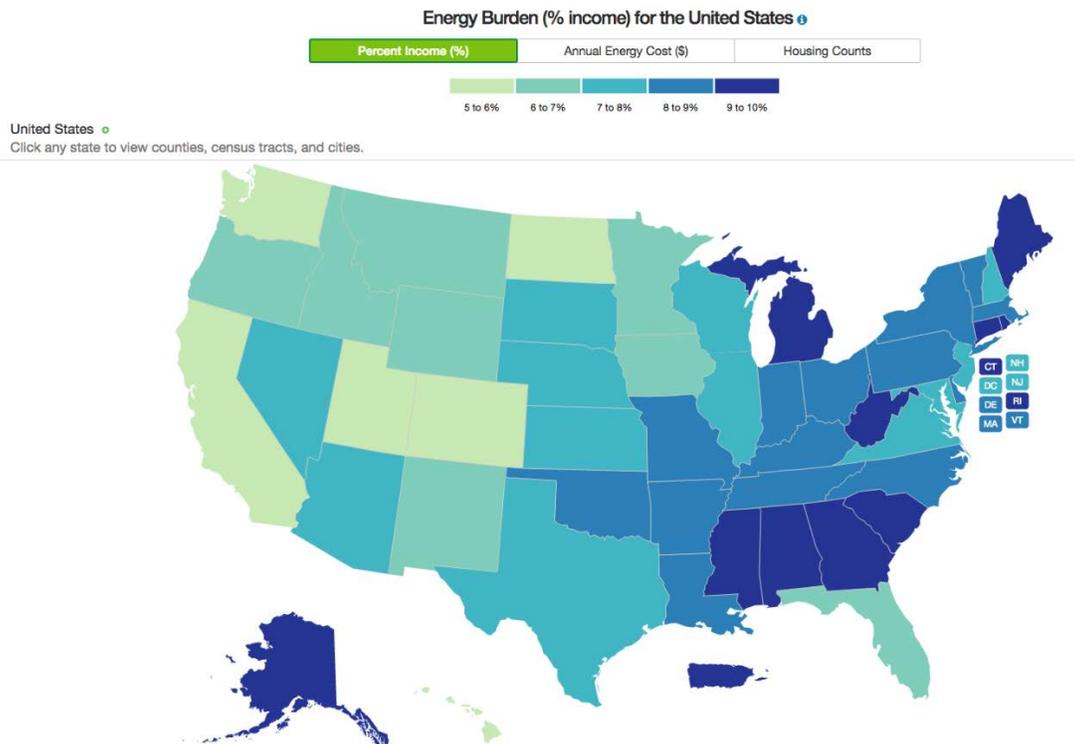


Figure 1: National choropleth map showing estimated average energy burden for 50 U.S. states, the District of Columbia, and Puerto Rico.

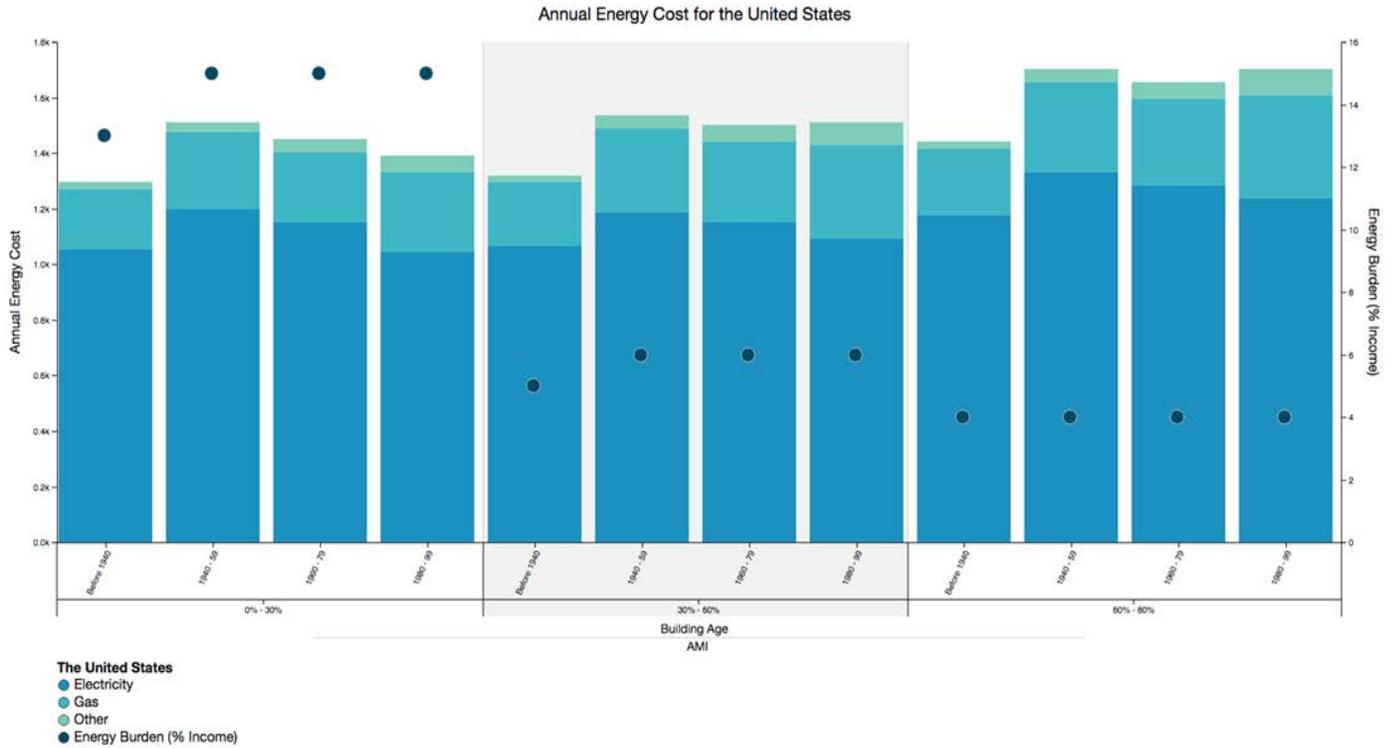


Figure 2: Breakdown of estimated average monthly housing energy costs by expenditure type (stacked bar, left axis) and the associated estimated average energy burden (dots, right axis) by building year of first construction and household income level as a percent of area median income.

The LEAD Tool includes both map-based visualizations and chart-based visualizations. Charts enable further disaggregation of the data by cohort. For instance, Figure 2 shows disaggregated monthly energy costs by energy type and by user-selected cohorts. In this case, cohorts are based on area median income and the building year of first construction. Other available cohort characteristics include tenure (owner or renter), number of units in the buildings (single family detached, single family attached, 2-unit, 3-4 unit, etc.), and housing unit primary heating fuel type (utility gas, bottled gas, electricity, fuel oil, etc.).

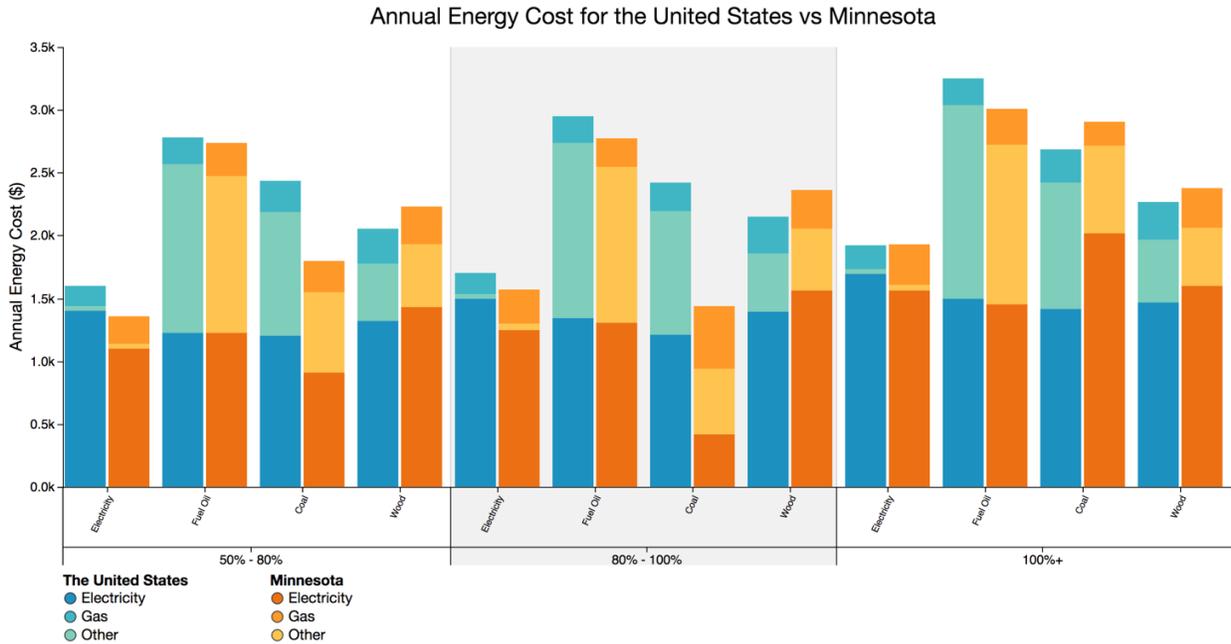


Figure 3: Estimated average annual energy costs for the U.S. (blue, left) and the state of Minnesota (orange, right) by housing unit primary heating fuel type and household income as a percent of area median income cohorts.

The chart-based visualizations also enable comparisons. For instance, Figure 3 shows estimated average energy costs for the entire U.S. compared with those for just the state of Minnesota. In this visualization, the user has selected area median income and heating fuel type to determine housing unit cohorts. All combinations of national, state, county, city, and tracts comparisons are possible.

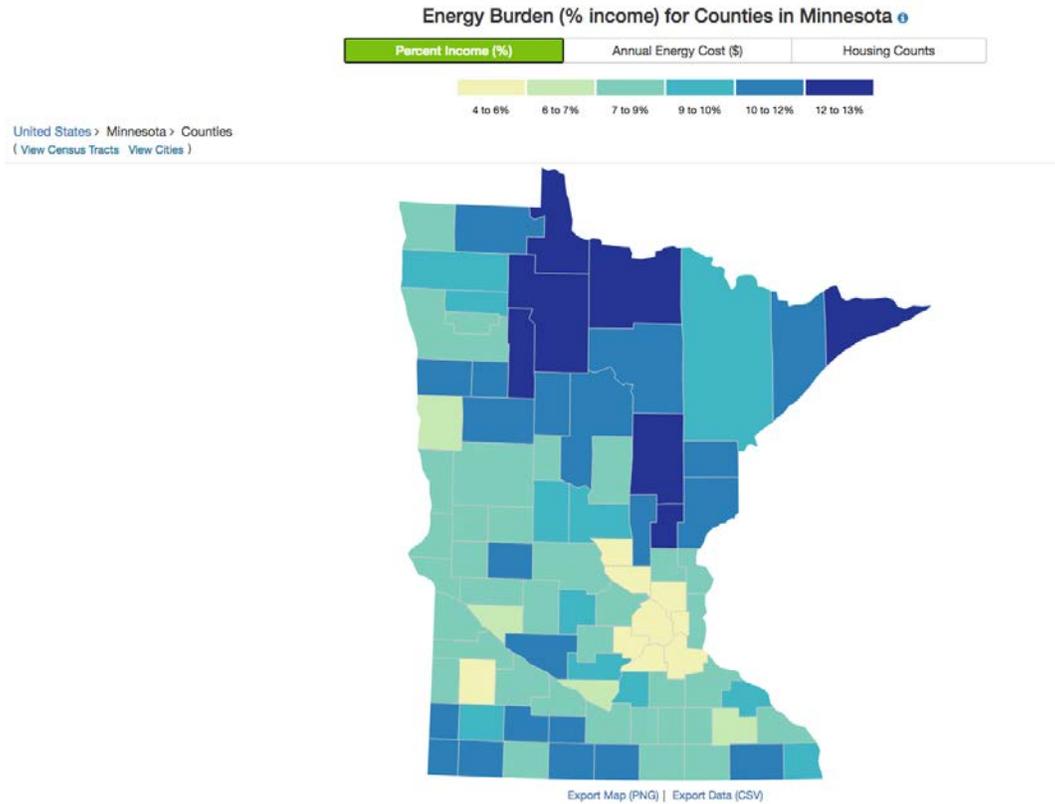


Figure 4: State choropleth map showing estimated average energy burden for all counties in Minnesota.

In addition to national and state level estimates and comparisons, the tool provides county, city, and tract-level options. For instance, Figure 4 is similar to Figure 1, but shows estimated average energy burden for counties across the state of Minnesota rather than states across the U.S.

3 Methodology

The LEAD Tool provides estimates of residential housing energy use including electricity, gas, and other fuels. The estimation process starts with the development of housing unit population distributions through the selection of cohorts assumed to have homogeneous energy use characteristics. In the residential housing sector, cohorts are based on locational, occupancy, physical, and demographic characteristics of housing units. Energy use values by cohort come from available statistical samples. These energy use values are then rescaled to match aggregate values from electric and natural gas utility reported sales and revenues. Finally, energy use values are calculated at the city and county levels. In this work, cities are generally defined as incorporated places. However, in several states, including New England states, New Jersey, New York, and Pennsylvania, minor civil divisions are also included in our definition of cities to capture coterminous towns and townships that may also serve as general-purpose local governments and perform the same governmental functions as incorporated places. In the state of Hawaii, which has only one incorporated place, counties are used in lieu of incorporated places. The detailed methodology is provided below.

3.1 Household Energy Estimates

Estimates of residential energy consumption are based on cross-tabulations of U.S. Census housing data from the 2016 5-year American Community Survey (ACS5).⁴ Average energy expenditures by different housing unit types are weighted by their housing unit counts to develop census tract-level estimates of energy expenditures. Those energy expenditure estimates are then rescaled to give energy consumption values, as described later. They are finally aggregated to cities and counties based on the census block-level occupied housing unit counts.

Spatial allocation of different housing unit types relies on the use of an iterative proportional fitting (IPF) algorithm.^{5,6} IPF is used sequentially to build increasingly complex cross-tabulations. Census tract-level published tables from the ACS5⁷ are used as the marginal totals, and cross-tabulations of the ACS5 Public Use Microdata Samples for the corresponding Public Use Microdata Areas are used as the seeds in the IPF algorithm. The resulting cross-tabulations (see Table 1) include housing unit tenure, building year of first construction, number of units in the building, primary heating fuel type, number of persons, and household income. Finally, household income and number of persons are collapsed to a single variable using either the U.S. Department of Housing and Urban Development definition of AMI⁸ or the U.S. Department of

⁴ U.S. Census Bureau. 2018. “American Community Survey Data.” <https://www.census.gov/programs-surveys/acs/data.html>.

⁵ IPF is applied in Lovelace, R. 2014. *Introducing Spatial Microsimulation with R: A Practical*. National Centre for Research Methods Working Paper. http://eprints.ncrm.ac.uk/3348/4/spat_microsimulation_R.pdf.

⁶ IPF is also introduced in Pitchard, D. R. & Miller, E. J. 2012. Advances in population synthesis: fitting many attributes per agent and fitting to household and person margins simultaneously. *Transportation* 39(3), 685–704. <http://www.springerlink.com/index/10.1007/s11116-011-9367-4>

⁷ U.S. Census Bureau. 2018. “American Community Survey Data.” <https://www.census.gov/programs-surveys/acs/data.html>.

⁸ U.S. Department of Housing and Urban Development. 2016, “FY2016 Income Limits.” <https://www.huduser.gov/portal/datasets/il.html#2016>.

Health and Human Services definition of FPL⁹. Note that we modify the published area median income cutoff of 50% to 60% by linearly interpolating between the published 50% and 80% income levels; for the State of Connecticut, we replace federal values for state values¹⁰; and for Puerto Rico, federal poverty level is the same as for the CONUS.

Table 1. Cross-Tabulation of ACS5 Data

Variable	Categories	ACS5 Published Table	ACS5 Microdata Sample
Tenure	Owner-Occupied and Renter-Occupied	Included below ^a	TEN
Building year of first construction	2010 and later, 2000-2009, 1980-1999, 1960-1979, 1940-1959, and 1939 and before	B25036, B25127 ^b	YBL
Number of units in the building	1 Unit Detached, 1 Unit Attached, 2 Units, 3-4 Units, 5-9 Units, 10-19 Units, 20-49 Units, 50 and More Units, and Mobile and Other Units	B25032, B25124 ^b , B25127 ^b	BLD
Primary heating fuel type	Utility Gas, Bottled Gas, Electricity, Fuel Oil, Wood, Coal, Solar, Other, and None	B25117	HFL
Number of persons	1-Person, 2-Person, 3-Person, 4-Person, 5-Person, 6-Person, 7 or More Persons	B25009, B25124 ^b	NP
Household income	0-5K, 5-10K, 10-15K, 15-20K, 20-25K, 25-35K, 35K-50K, 50-75K, 75-100K, 100-150K, 150K and more	B25118	HINCP
Area median income	0-30%, 30-60%, 60-80%, 80-100%, greater than 100%	not available	not available
Federal poverty level	0-100%, 100-150%, 150-200%, 200-400%, greater than 400%	not available	not available

^a Many of the published tables in the ACS5 are broken out by tenure.

^b The ACS5 published tables include some cross-tabulations, including building year by number of units and number of persons by number of units. These cross-tabulations are incorporated in the IPF sequence to improve agreement between housing unit estimates and published Census values.

For each Public Use Microdata Area, corresponding energy expenditure values are calculated for each housing unit type in the above cross-tabulation. This requires modification to the microdata samples. To improve data accuracy, we removed survey responses for which housing energy costs are included in other housing costs, or where energy costs across multiple fuels type are combined. Census tract-level housing unit counts are used to develop weighted averages.

To validate estimates, additional steps were taken to separate bottled and utility gas use. The ACS5 provides energy expenditures for three categories of heating fuel types: electricity, gas, and other fuels. We assume that electricity is always provided by a utility. However, this assumption does not hold true for gas, which includes both utility-delivered natural gas and various forms of bottled gas. To make separate estimates of utility and bottled gas, we assume that housing units reporting bottled gas as their primary heating fuel type do not have access to

⁹ US Department of Health and Human Services. 2016. “Computations for the 2016 Poverty Guidelines.” <https://aspe.hhs.gov/computations-2016-poverty-guidelines>

¹⁰ Connecticut State Department of Housing. 2016. “2016 Combined Income Limits.” <https://portal.ct.gov/DOH/DOH/Additional-program-pages/HUD-Rent-and-Income-Limits>.

utility gas. Those gas expenditures are subtracted from the associated census tract-level values to yield estimates of only utility gas expenditures.

As demonstrated in Section 5 Validation of Estimates to Measured Values, resulting values are in good agreement with utility-reported values. Nevertheless, this approach has several shortcomings aside from the more general issues associated with self-reported survey data. First, electricity and gas expenditures are taken for only a single month and that month is not reported publicly. Given the strong seasonal variation of energy use, extrapolation of annual values from monthly values hinges upon the quality of the sampling to cover all months of the year. This cannot be verified with the public data. Second, the ACS5 includes only occupied housing units. Unoccupied housing units may consume a significant fraction of energy. Third, elimination of responses reporting no expenditures associated with fuel use introduces unknown biases into the estimates. Fourth, the availability of piped utility natural gas is unknown and must be inferred from the prevalence of utility gas as a heating fuel source.¹¹ Finally, the household energy characteristics of census tracts may differ significantly from their larger Public Use Microdata Area (PUMA), leading to poor IPF estimation. (For simple comparison, there are roughly 4,500 PUMAs and 73,000 census tracts in the U.S.)

¹¹ In some cases, the ACS5 incorrectly assigns utility gas heated housing units to areas lacking utility gas availability, which can lead to errors in this modeled data.

4 Calibration

Modeling energy consumption and expenditures at the census tract level represents a bottom up approach. Estimates can be improved through fully leveraging existing aggregate information as well as the data available at different geographies and cohorts used to build the bottom up approach. To combine the bottom up and top-down approach and create a more robust model, we rescale census-tract estimates using energy intensity weighting and sum to state electricity and natural gas sales totals reported in Energy Information Administration Forms 861 and 176.

This rescaling or calibration process of allocating total state energy consumption to each census tract uses two key assumptions. For electricity, it was assumed that all tracts are served by a utility company. This assumption is supported by the high electrification rate found in the United States. For natural gas, it was assumed that not all tracts are served by utility gas, as other fuel types such as bottled gas, coal, or liquid fuel may serve as substitutes for utility gas.

To inform the subsequent steps, electric and natural gas utility territories were mapped to census tracts. This mapping was conducted using the Ventyx utility territories spatial data, which provides boundary information for electric and natural gas companies.¹² A simple intersection was conducted between the territories and the census tracts. Utility IDs were also normalized to facilitate merging the spatial representation of the territories and the data provided by the Energy Information Administration.

Mapping utility service territories to census tracts sought to achieve two goals. First, it was important to know which tracts should be excluded from the natural gas allocation process. Second, to rescale bottom up estimates using energy intensity weighting, each tract was assigned the energy sales rate (dollars per megawatt-hour for electricity and dollars per million cubic feet for natural gas) of the utility that serves it, which was obtained from the Energy Information Administration's data files. For tracts that had no electric utility initially assigned to it, as in states with unbundled utility service, the state average per-unit commodity and delivery service values price was used. For tracts where several utilities were present, the average of the prices was used.

Census tract-level consumption and expenditure estimates were geographically rescaled to match utility sales (electricity) or volumes (natural gas) and revenues. A weighting approach was then applied.

4.1 Energy intensity weighting

Rescaling cannot be accomplished exactly for the residential sector because the number of utility customers does not correspond to the number of housing units, and utility territories often overlap. Therefore, the calibration process first assigns utility customers by utility to census tracts using the IPF algorithm. The geospatial overlay of Ventyx utility boundaries with census tracts yields the binary matrix \mathbf{U} of 1's and 0's, with columns corresponding to utilities and rows

¹² ABB/Ventyx Energy Velocity Suite. 2016. Local Distribution Company Territories. Retrieved June 1, 2018.

to census tracts. Column marginal sums are utility customers and row marginal sums are total housing units, buildings, and establishments. The resulting matrix \mathbf{C} is the approximate assignment. Note that the IPF does not converge as the column and row marginal sums total to different values. However, each iteration preserves the corresponding marginal sum, in this case, chosen to be utility customers to match the Energy Information Administration data.

Per customer consumption and expenditures are assumed to be similar among customers within a single census tract but served by different utilities. Based on this assumption, per customer consumption \mathbf{l} and per customer expenditures \mathbf{l}' should also be similar whether taken from the utility-reported values, with \mathbf{S}_{rep} as a diagonal matrix of utility-reported sales and volumes per customer,

$$\mathbf{l} = [\mathbf{U} \cdot (\mathbf{C} \cdot \mathbf{S}_{rep})^T] \cdot \mathbf{1}$$

or estimated from the customer perspective, with \mathbf{R}_{est} as a diagonal matrix of estimated consumption per customer.

$$\mathbf{l}' = [\mathbf{U} \cdot (\mathbf{C}^T \cdot \mathbf{R}_{est})] \cdot \mathbf{1}.$$

Matrix multiplication by the column vector of ones $\mathbf{1}$ yields a row-wise sum. The estimated values are then rescaled by the ratio of these quantities to give the following calibrated result

$$\mathbf{R}_{cal} = \mathbf{R}_{est} \cdot (\mathbf{l} \oslash \mathbf{l}')$$

where the operator \oslash represents element-wise division. Derivation of these equations is provided in Appendix A. The diagonal elements of \mathbf{R}_{cal} are the calibrated census tract-level estimated consumption per customer.

The calibration approach described in Appendix A leads to improved agreement with available measured inventory data. However, there are several instances where this approach can lead to poor agreement:

- The equivalence of \mathbf{l} and \mathbf{l}' holds under the assumption that census tracts are relatively homogeneous in per-customer energy consumption. However, the calibration process is biased toward larger utilities in instances where two or more utilities share a single census tract and those utilities have very different per customer energy consumption values.
- There are data gaps and inaccuracies in the mapping of utilities to census tracts, leading to incorrect assignment of customers to utilities and customers to census tracts.
- The IPF approach may poorly reflect the true geographic distribution of utility customers.
- Utility service territories may be large, encompassing many cities and counties. For instance, in a severe case, most of the state of Rhode Island is served by a single distribution utility. In this case, the calibration process merely reflects the state averages and provides poor results for areas served by the state's two much smaller municipal utilities.

5 Validation of Estimates to Measured Values

5.1 Validation of Modeled Data

There are several publicly available sources of data for validation purposes. Xcel Energy, an investor-owned utility that operates in several western and midwestern states, reports community-level electricity and natural gas sales and revenue data.¹³ Also, many cities in Minnesota report consumption data through the Regional Indicators Initiative.¹⁴ Further, Massachusetts reports community-level electricity data through Mass Save.¹⁵ Finally, many local governments have undertaken greenhouse gas emissions inventories. As part of an intermediary step, those inventories often report electricity and natural gas consumption by sector.

One significant difficulty with validation is that the number of reported utility customers may differ significantly from the number of occupied housing units. There are several reasons why this occurs. Most commonly, there are discrepancies between the community boundaries used by the reporting entity and the community boundaries used by the U.S. Census. There are also difficulties with estimating the number of utility customers. While we can assume that most occupied housing units are served by an electric utility, there may be many unoccupied housing units that also are served by an electric utility, and multi-family units may share a single meter. Furthermore, many occupied housing units are not served by a natural gas utility, and multi-family units may share a single meter or common heating unit.

Our assignment of natural gas customers to census tracts is approximate. Comparisons of state- and local-calibrated estimates to community data from Xcel Energy (electricity and natural gas) and Mass Save (electricity only) are shown in Figures 5 and 6, where each community-wide consumption data point reported is charted according to its difference with modeled data. State-level calibration rescales energy consumption estimates to match state-wide totals (see Figure 5). However, local-calibrated estimates (see Figure 6) rescale energy consumption based on utility-level totals (as described in Section 3.1 and Appendix A).

¹³ Xcel Energy. 2018. “Community Energy Reports.” https://www.xcelenergy.com/working_with_us/municipalities/community_energy_reports.

¹⁴ Regional Indicators Initiative. n.d. “Measuring City-Wide Performance.” <https://www.regionalindicatorsmn.com/>.

¹⁵ Mass Save. 2019. “Mass Save Homepage.” <https://www.masssave.com>

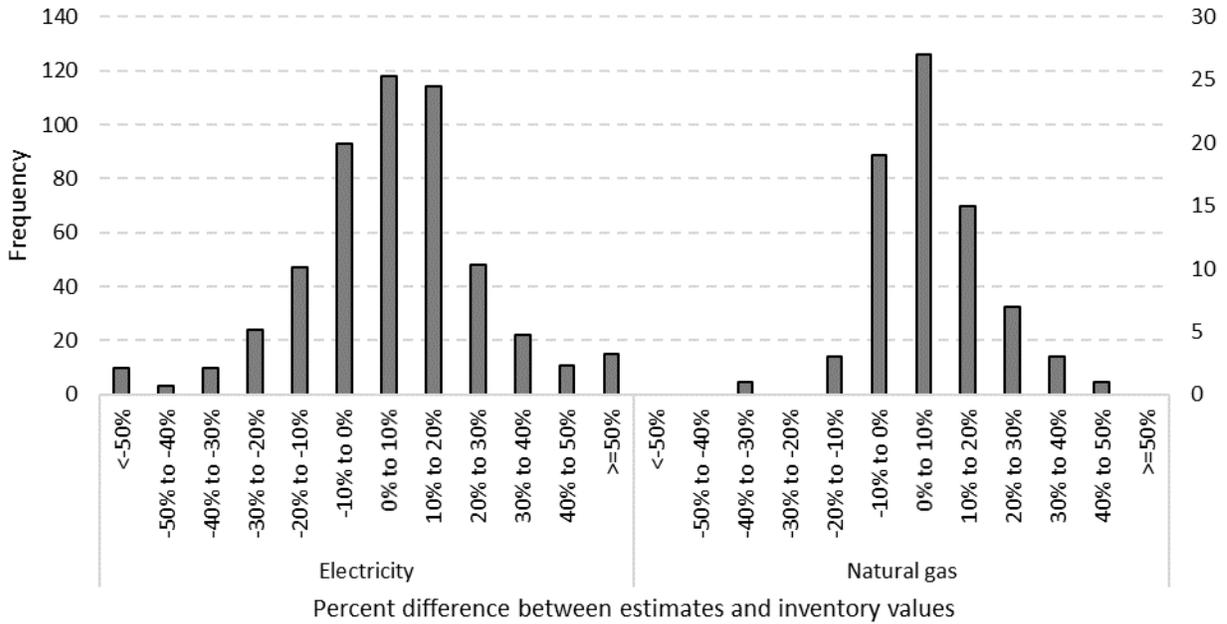


Figure 5. Comparison of state-calibrated residential estimates to community data from Xcel Energy and Mass Save.

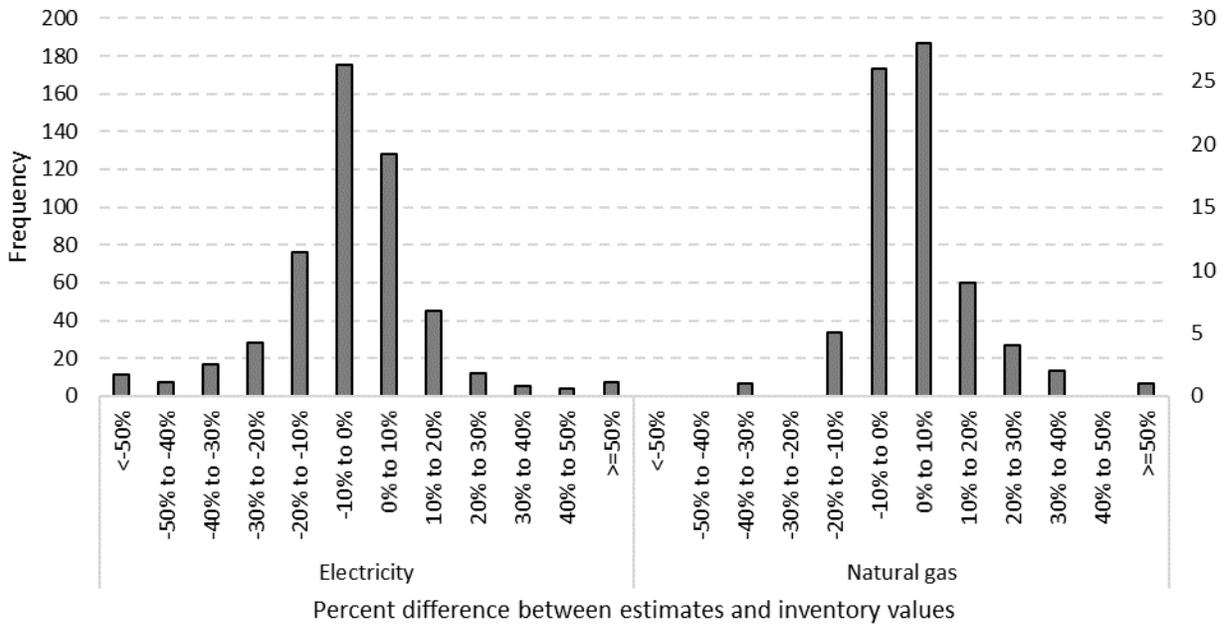


Figure 6. Comparison of local-calibrated residential estimates to community data from Xcel Energy and Mass Save.

This comparison illustrates that the majority of residential energy consumption estimates based on the methodology described herein are within +/- 10% of the 515 reported electricity and 76 reported natural gas city and county inventory values provided by Xcel Energy and Mass Save. There are several outliers with large discrepancies. Many of those communities are very small,

with less than 1000 occupied housing units, where potential errors in the Census survey data or differences in geographic boundary definitions between utilities and Census could generate large differences on a percentage basis. To add, the Mass Save data does not include customer counts, preventing basic checks against this issue.

6 Discussion

The Low-income Energy Affordability Data (LEAD) Tool seeks to address current gaps in residential housing energy data. Most energy data are available only at the aggregate level. However, strategic development and implementation of energy programs and policies, such as those targeting low-income populations, require disaggregated information on the specific subpopulations of interest. Even in cases where stakeholders have access to household utility data, there is limited information regarding the characteristics of those households such as income level or housing unit type. Given the absence of measured data, the LEAD Tool employs estimation techniques based on widely available statistical data from the U.S. Census Bureau and the Energy Information Administration. This approach has the advantage of providing national coverage for cities and counties across all 50 U.S. states, D.C., and Puerto Rico.

Electricity consumption estimates are within +/-10% for 59% of the 515 community reported values. The agreement increases to 73% if considering the 92 larger communities with more than 20,000 occupied housing units. For natural gas consumption, those percentages are 71% for 176 communities and 88% for the subset of 34 larger communities. While comparisons at further levels of disaggregation, for instance by income level or housing unit type, are unavailable, these aggregate comparisons lend confidence to the validity of the estimation approach. Overall, the approach results in a comprehensive, demographically and geographically resolved, and standardized dataset.

While the LEAD Tool provides demographic and spatial granularity, due to the nature of the underlying statistical data and estimation approach, it does not provide temporal granularity. The ACS data comes from a 5-year rolling average which smooths annual changes. Furthermore, the ACS and our estimation approach result in uncertainties and margins of error. While these have not been calculated explicitly, we expect them to exceed any apparent differences from one LEAD Tool vintage to the next.

The web-based LEAD Tool allows users to quickly and easily develop residential housing energy profiles. These profiles provide them information such as:

- How the need for weatherization and/or low-income energy assistance differs across geographies based on the prevalence of those households and the relative magnitudes of their average energy burdens,
- How the housing stock differs among low, moderate, and high income households and the relative need for different types of energy efficiency strategies,
- How changes to retail energy costs could differentially impact the energy burdens of low, moderate, and high income households,
- How energy bill savings, through various energy efficiency programs, could impact energy burden for low-income households, and
- How those with high electricity energy bills can be a potential participant for renewable energy, if cost is comparably lower for renewable energy than electricity

Going forward, additional work is needed to improve agreement with measured values. This requires both increasing the number of comparisons to reported values, for instance through data scraping of greenhouse gas emissions inventories, as well as analyzing the sources of error. There may be systematic drivers for discrepancies and the identification of those drivers may lead to improved methodologies. Furthermore, a weather-based approach may also provide better estimates, with the added benefit of further disaggregating energy expenditures into heating, cooling, and baseload end-use demands.

Lastly, there may be opportunities to supplement measured data with the LEAD Tool, in order to fill informational gaps and improve calibration of estimates. For instance, Weatherization Assistance Program, Low Income Home Energy Assistance Program providers, and the state agencies that oversee those programs may have access to information on applicants and participants.¹⁶ These data points, while representing measured data rather than estimates, are not statistical samples; and conclusions based on those data points may have unknown biases. Putting those data points in the context of the full low-income household populations from the LEAD Tool may reveal gaps in program coverage or design.

¹⁶ North Carolina Clean Energy Technology Center and Upper Coastal Plain Council of Governments. (2019). “Powering Energy Efficiency and Impacts: A Data-Driven Project Supporting Low-Income Households in Northeastern North Carolina.” <https://www.energy.gov/eere/analysis/powering-energy-efficiency-and-impacts-project-framework-peeif>.

Appendix A. Derivation of Calibration Equations

The calibration approach seeks to identify groupings of census tracts within states that share a set of utilities. Then, the average per customer consumption over those census tracts should equal the corresponding average per-customer sales by the associated utilities. This assumes that states are separable in the sense that there are smaller than state-level groupings.

Start with a matrix \mathbf{C} , which maps utility customers (columns) to census tracts (rows), and a matrix \mathbf{U} , which is a binary representation of \mathbf{C} :

$$U_{ij} = \begin{cases} 1 & \text{if } C_{ij} \neq 0 \\ 0 & \text{if } C_{ij} = 0 \end{cases} \quad \text{Eq. A.1}$$

Then, the following quantity provides the total number of utility customers shared by each census tract

$$\mathbf{c} = [\mathbf{U} \cdot \mathbf{C}^T] \cdot \mathbf{1} \quad \text{Eq. A.2}$$

where matrix multiplication by the column vector of ones $\mathbf{1}$ yields a row-wise sum. For example, suppose there are four census tracts labeled 1, 2, 3, and 4 and three utilities labeled A, B, and C, as illustrated in Figure A1.

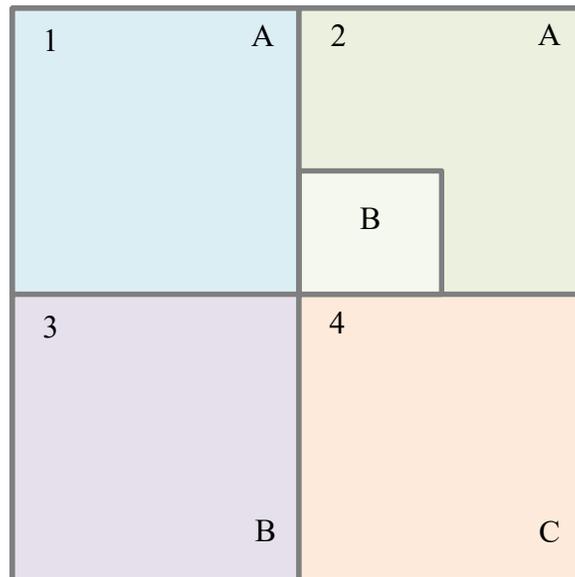


Figure A1. Example census tracts 1, 2, 3, and 4 and utility service areas A, B, and C

Further suppose this corresponds with the following mapping:

$$\mathbf{C} = \begin{bmatrix} 400 & 0 & 0 \\ 200 & 50 & 0 \\ 0 & 300 & 0 \\ 0 & 0 & 600 \end{bmatrix}$$

which assigns 400 customers of utility A to census tract 1, 50 customers of utility B to census tract 2, and so forth. Then, Eq. A.2 gives:

$$\mathbf{c} = \begin{bmatrix} 600 \\ 950 \\ 350 \\ 600 \end{bmatrix}$$

which shows that that census tract 1 shares 200 utilities customers in census tract 2 to give a total of 600 customers, census tract 2 shares 400 customers in census tract 1 and 300 customers in census tract 2 to give a total of 950 customers, and so forth.

The average per customer sales for each grouping may be calculated accordingly:

$$\mathbf{s} = \left\{ \left[\mathbf{U} \cdot (\mathbf{C} \cdot \mathbf{S}_{rep})^T \right] \cdot \mathbf{1} \right\} \oslash \left\{ \left[\mathbf{U} \cdot \mathbf{C}^T \right] \cdot \mathbf{1} \right\} \quad \text{Eq. A.3}$$

where the operator \oslash represent element-wise division. Here, we have simply inserted \mathbf{S}_{rep} as a diagonal matrix of utility reported sales and volumes per customer to produce a customer-weighted average.

This same quantity may be calculated based on consumption per customer at the county level. In this case, the numerator is revised to read:

$$\mathbf{r} = \left\{ \left[\mathbf{U} \cdot (\mathbf{C}^T \cdot \mathbf{R}_{est}) \right] \cdot \mathbf{1} \right\} \oslash \left\{ \left[\mathbf{U} \cdot \mathbf{C}^T \right] \cdot \mathbf{1} \right\} \quad \text{Eq. A.4}$$

where \mathbf{R}_{est} is a diagonal matrix of estimated consumption per customer.

Under the assumption that utility customers of different utilities but within a single census tract have the same average per-customer consumption, the equality $\mathbf{r} = \mathbf{s}$ should hold. Then, we can try to enforce the equality accordingly:

$$\mathbf{R}_{cal} = \mathbf{R}_{est} \cdot (\mathbf{s} \oslash \mathbf{r}) \quad \text{Eq. A.5}$$

where the diagonal elements of \mathbf{R}_{cal} are the calibrated census tract-level estimated consumption per customer. This equation is equivalent to equation Y if we simply set the numerator of \mathbf{s} as \mathbf{l} and the number of \mathbf{r} as \mathbf{l}' .

To complete the simple example, if we assume:

$$\mathbf{S}_{rep} = \begin{bmatrix} 14 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 16 \end{bmatrix}$$

and

$$\mathbf{R}_{est} = \begin{bmatrix} 2300 & 0 & 0 & 0 \\ 0 & 2100 & 0 & 0 \\ 0 & 0 & 1700 & 0 \\ 0 & 0 & 0 & 2700 \end{bmatrix}$$

then

$$\mathbf{s} = \begin{bmatrix} 14 \\ 12.5 \\ 10 \\ 16 \end{bmatrix}$$

and

$$\mathbf{r} = \begin{bmatrix} 2233 \\ 2058 \\ 1757 \\ 2700 \end{bmatrix}$$

The calibrated per-customer consumption values are then:

$$\begin{aligned} \mathbf{R}_{cal} &= \begin{bmatrix} 2300 & 0 & 0 & 0 \\ 0 & 2100 & 0 & 0 \\ 0 & 0 & 1700 & 0 \\ 0 & 0 & 0 & 2700 \end{bmatrix} \cdot \left(\begin{bmatrix} 14 \\ 12.5 \\ 10 \\ 16 \end{bmatrix} \oslash \begin{bmatrix} 2233 \\ 2058 \\ 1757 \\ 2700 \end{bmatrix} \right) \\ &= \begin{bmatrix} 14.4 & 0 & 0 & 0 \\ 0 & 12.8 & 0 & 0 \\ 0 & 0 & 9.7 & 0 \\ 0 & 0 & 0 & 16 \end{bmatrix} \end{aligned}$$

We can check this numerical result by re-calculating the utility values:

$$\mathbf{S}_{est} = \{\mathbf{C} \cdot [\mathbf{1} \cdot \mathbf{R}_{cal}]\} \oslash \{\mathbf{1} \cdot \mathbf{C}\} = [13.9 \quad 10.1 \quad 16] \quad \mathbf{Eq. A.6}$$

showing good agreement with the diagonal values in \mathbf{S}_{rep} .

In this example, we have provided numerical values of similar magnitudes to the values used in the residential calibration. A value of 14 corresponds to 14 MWh/year, which is within the normal range of annual household electricity consumption. A value of 2,300 corresponds to \$2,300/year in electricity expenditures, also within the normal range. Residential estimates come from the 2016 ACS5, which samples expenditures and not consumption. However, this analysis assumes that expenditures are roughly proportional to consumption.

Note that the choice of \mathbf{R}_{cal} is not unique, and in general, there are many possible values that provide agreement with the utility-reported values. However, this approach has the benefit of making a choice that retains some sense of the original guess \mathbf{R}_{est} , as most choices of \mathbf{R}_{cal} are

not physically plausible. One could further refine this approach by taking the calibrated value as the seed value in a numerical minimization, such that the utility values are matched exactly. Minimizing the difference between the values in Eq. A.6 and the diagonal elements of Eq.A.5 leads to the following result. Differences are less than 2%.

$$\mathbf{R}_{fit} = \begin{bmatrix} 14.6 & 0 & 0 & 0 \\ 0 & 12.8 & 0 & 0 \\ 0 & 0 & 9.5 & 0 \\ 0 & 0 & 0 & 16 \end{bmatrix}$$

Improving agreement with utility values is desirable; however, in practice, performing this minimization step is challenging. Any errors or unusually divergent values in \mathbf{C} and \mathbf{S}_{rep} can lead to \mathbf{R}_{fit} values that deviate significantly from the original \mathbf{R}_{est} guesses. Thus, in this report, we omit this second step and limit the calibration to Eq. A.5.