HOME ENERGY CONSUMPTION ESTIMATION BY END USE AND ENERGY EFFICIENCY UPGRADE RECOMMENDATIONS

by

Chao Li

Dr. Peter Haff, Advisor

May 2014

Master project submitted in partial fulfillment of the

requirements for the Master of Environmental Management degree in

the Nicholas School of the Environment of

Duke University

2014

____

MP Advisor's signature

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Today, residential energy consumption is becoming an ever-growing sector, world widely representing 16-50% (30% average) of total energy consumed by all other sectors (Saidur, Masjuki, and Jamaluddin 2007). In the United States, residential sector approximately accounts for 21.2% of total energy consumption of year 2012 (EIA 2014). These data indicates the crucial role that residential sector plays in total energy consumption, meaning a necessity of getting a deep understanding of the associated energy consumption characteristics to better prepare for the increasing energy demand in the future.

Motivations for residential energy efficiency upgrades

Considering the significant proportion of residential energy consumption, potential energy efficiency upgrades that are targeted in residential sector could reduce the associated energy usage at a considerable level. In Annual Energy Outlook (AEO) 2013, Energy Information Administration (EIA) projected that in 2040, even with increased residential energy demand, efficiency improved by different cases can eventually offset the demand growth and reduce residential energy consumption, as shown in Figure 1 (EIA 2013a).

Before implementing energy efficiency retrofits, it is crucial to obtain detailed information on breakdown of residential energy use. For instance, Heating, Ventilation and Air Conditioning (HAVC) system of higher efficiency could potentially save 30% energy in space-heating/cooling by replacing old HAVC system (Jia et al. 2012). But without the end energy use estimate (i.e. energy consumed by spacing heating/cooling), it is not wise to blindly replace the existing HAVC system, given the possibility that the heating/cooling sector may only accounts for a tiny part of total consumed energy. Hence, in an effort to better implement targeted energy efficiency improvement strategy, it is essential to understand the detailed information on breakdown of residential energy use, which in this project, is specified into 4 different end use categories: space heating, water heating, water cooling, and appliances.

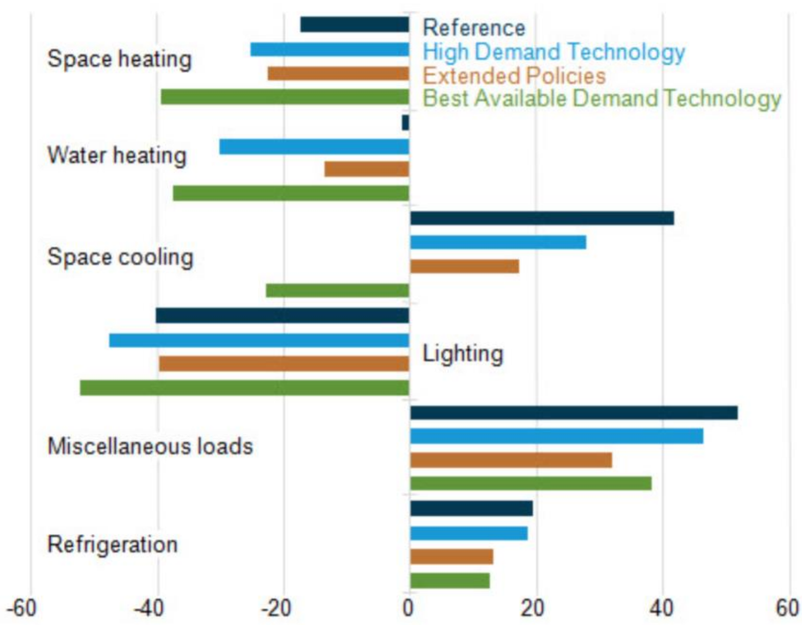

Figure 1: Change in residential delivered energy consumption for selected end uses in four classes, 2011-2040 (percent)[1]

Variations of residential energy use characteristics

A variety of variations exist in residential energy use characteristics, such as occupant behavioral pattern, and efficiency standards of equipment, making it challenging to accurately estimate the breakdown of energy use. A previous research explored both qualitative and quantitative effects of occupancy and behavioral on residential energy use (Seryak and Kissock 2003), which focused on the residential homes owned by University of Dayton (UD). For the electricity use, influence factors include:

- Number of occupants: Electricity use is positively correlated with the number of occupants per household, that is, the greater number of occupants a household has, the higher the electricity use is. But electricity consumption per capita tends to decrease sharply as the number of

[1] *Retrieved from AEO 2013, EIA. 'The High Demand Technology and Best Available Demand Technology cases assume different levels of efficiency improvement without anticipating new appliances standards. The Extended Policies case assumes the enactment of new rounds of standard, generally based on improvements seen in current ENERGY STAR equipment.' In the reference case, the number of households increases by 32 percent, together with a 41 percent increase in square footage from 2011 to 2040, without efficiency improvements.*

occupants increases.

- Time of occupancy: Electricity consumption differs and is un-evenly distributed in each period of a year. Electricity use tends to peak at summer and winter because of the great demand for space cooling and heating; households near college usually have lower energy consumption in June due to the summer break, while they also face a sharp increase in September, when students start classes.
- Occupant behavior: Even in the situation where some of the energy use characteristics are identical in two households, other variations may still lead to a big difference in electricity use, among which occupant behavior plays a crucial role in controlling the electricity bill. Occupant behaviors depend on various factors, ranging from local temperature to regional energy policies. For instance, residents in Florida are less likely to consume more energy in space heating than those from Minnesota. The research found that with the same number of occupants and occupancy periods, electricity consumption still changes at a great level due to the variation of occupant behaviors.

Current obstacles in achieving breakdown information

In addition to the discussed energy consumption characteristics above, energy use is decided by other various parameters. Though realizing those variables is necessary to estimate the breakdown of energy use, it is hindered by difficulties in quantifying their values, such as behavioral patterns, an abstract term that is hard to be quantified.

Other major sectors such as transportation, industrial, and commercial, are well understood due to more centralized ownership, self-interest in cutting energy consumption, and well-documented regulations. However, detailed breakdown on residential sector is limited by several reasons (Swan and Ugursal 2009):

- Wide variation of structure sizes, geometrics, and thermal envelope materials.
- Energy consumption patterns vary in different household occupant.
- Privacy issues limit energy-related household data collection.
- Prohibitive cost of detailed sub-metering of household end uses.

Methodologies of residential energy use modeling

Complexity of residential energy use patterns and dependence on data input level make modeling residential energy use potentially challenging. However, based on the different capabilities, strengths, weaknesses, and applicability of each modeling technique, matching input data with models that can best use them could produce satisfactory models. Generally speaking, techniques employed to model residential energy use can be classified into two categories, “top-down” and “bottom-up”, and this terminology is referred to the hierarchal position of data inputs (Swan and Ugursal 2009), as indicated in Figure 2.

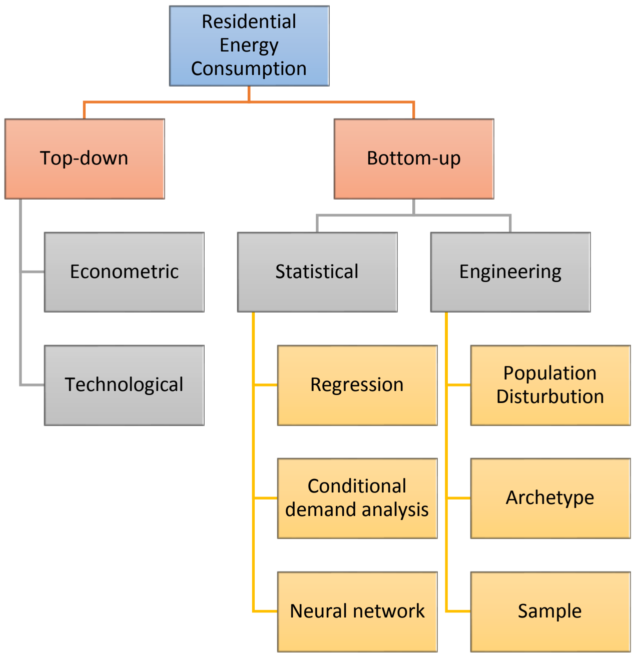

Figure 2. Modeling techniques for estimating the regional and national residential energy consumption[2]

Top-down approach

This approach evaluates the residential energy sector with a special emphasis on the effects of long-terms changes, that is, rather than focusing on the effects from individual energy consumption, it suggests considering the residential sector as an energy sink, and using estimates of total residential sector energy consumption as well as macro-variables, such as macroeconomic indicators, climatic

[2] *Reproduced from Swan, Lukas G., and V. Ismet Ugursal. 2009. “Modeling of End-Use Energy Consumption in the Residential Sector: A Review of Modeling Techniques.”*

conditions, and demolition rate in the residential sector, to attribute the energy consumption to characteristics in the whole residential sector. For example, if demolition rate increases by 5%, a top-down model may estimate that the residential sector could consume 2% less energy, due to a decreased number of occupants.

One primary advantage of top-down model is the data accessibility. As mentioned above, it employs many commonly used variables, many of which are of great availability and are from historical dataset. But, the reliance on historical dataset is also a disadvantage of top-down model, because of its incapability to model discontinuous technology advances that could significantly influence a variable by a technology breakthrough. Additionally, omitted information on individual energy end uses further keeps top-down model from reducing the energy consumption, due to its incapability of identifying key sectors of potential energy efficiency improvements (Swan and Ugursal 2009).

Bottom-up approach

Contrary to top-down approach, bottom-up approach includes models that utilize input data of micro-level and account for the energy consumption from individual end-use. Rather than using the macro indicators to forecast or estimate the residential energy sector, bottom-up models are extrapolated through aggregating the individual energy consumption estimates to represent a certain region or even a nation.

In particular, bottom-up approach can be further classified into two categories, statistical model and engineering model. Engineering model can estimate energy consumption of various end uses through involving equipment usage and energy ratings, while statistical model can build a relationship between household's characteristics (e.g. income, householder's age) and different end uses, based on regression model and controlling for exogenous variables (e.g. climate conditions and household occupancy) (Min, Hausfather, and Lin 2010). Commonly used input data for bottom-up model has a wide scope, ranging from regional characteristics (e.g. household location, average temperature in winter, and regional geometry) to individual information (e.g. preferred indoor temperature, energy use patterns of occupant, and occupancy time). This high level of detail of input data makes bottom-up model capable of identifying areas of potential energy efficiency improvements, with the help from established relationship between household's characteristics and end uses. Also it can help determine the total energy consumption of residential sector without historical data, compared to up-down approach. The strength obtained by the high level of detailed input, however, also imposes a major drawback for bottom-up model, which are the lower data availability and associated model complexity caused by the various input data categories (Swan and Ugursal 2009). Particularly, in addition to the extensive data input requirement, engineering model is incapable of capturing variations due to socioeconomic-driven behavioral characteristic, further limiting their application in home energy end uses estimation, on which individual behavioral patterns usually have great influence.

Residential Energy Consumption Survey (RECS)

RECS is designed and administered by EIA, to collect energy consumption characteristics on individual housing unit, use patterns, and household demographics. Together with energy information from energy suppliers, survey data is further used to estimate energy usage of these surveyed households, by different end uses including space heating, cooling, water heating and other end uses (EIA n.d.). It is currently the only nationwide data source that provides detailed residential energy end uses by interviewing representative sample households. However, representative samples in RECS is selected under a multi-stage area probability design[3] (EIA 2009), leading to a limitation in analyzing special energy use patterns due to the lack of special characteristics other than census regions. Still, RECS is useful to explore how household characteristics, such as annual income, householder's age, and behavioral patterns, influence the energy use distribution to each energy end use (Min et al. 2010).

With the intention to obtain comprehensive energy consumption information from both consumer and supplier, RECS is designed to consist of three surveys, the Household Survey, the Rental Agency Survey, and the Energy Supplier Survey. The first two surveys are intended to collect energy consumption characteristics from selected housing units, while Energy Supplier Survey is designed to obtain specific energy consumption data of a certain housing unit sample, which is responded by energy companies that serve sampled housing units. Modeling the data collected from those surveys, EIA produces estimates for several home energy end uses, including space heating, cooling, water heating, and other sectors, all of which serve as dependent variable in this project.

---

[3] *Based on this probability-design sampling, RECS selects housing unit sample through the following stages: randomly choosing counties; randomly drawing segments that are generated by sub-dividing the selected counties; choosing a list of housing units for each segment; final housing unit samples are then drawn from the housing unit list.*

Statistical model approach

In this project statistical model approach is employed to establish the relationship between household's characteristics and various end uses (i.e. space heating, space cooling, water heating, and appliances), as one of the major objectives is to analyze the energy use breakdown for individual household. Specifically, multiple regression model methodology is introduced and implemented by using the variables and associated data obtained from Residential Energy Consumption Survey (RECS), and each energy end use has a distinct or similar model due to the variations existing in energy consumption characteristics. The established model can be further used to predict regional residential energy consumption by end uses. In this project, input data are primarily retrieved from U.S. Census, which has a database of zip-code level resolution and variables that are identical to those from RECS.

OBJECTIVES

The primary objectives of this project are:

- To establish multi-regression statistical models for home energy end uses estimation, including water heating, space heating, space cooling, and appliances. During the model establishment process, variables that are closely associated with energy consumption are introduced to better account for different energy end uses.
- Using the established statistical models, to deliver an accurate breakdown of energy consumption for a specific individual, given a personalized input for variables of regression models.
- To produce a zip-code level database of household energy use breakdown (Durham is used as an example in this project), using the established statistical models and data from U.S. Census. In the absence of personalized input information, model itself cannot provide any energy breakdown estimation, but the U.S. census data can be employed as the variable input to produce a generalized estimation for a certain region, which can serve as a reference with the deficiency of personalized information.
- To offer targeted recommendations for each end use based on the estimates, aiming to effectively improve energy efficiency.

METHODOLOGY

Overview

In this project, statistical model that regresses on multiple variables, chosen from RECS, is employed to estimate home energy consumption by end use. Using the built regression models, dataset containing generalized home energy end uses for a specific location (e.g. Durham) are produced through importing data from U.S. Census, which can achieve estimates of zip-code resolution. In addition, an excel model is also built to produce estimates of higher accuracy, given a personalized input data that corresponds to the variables in multi-regression model. The following sections describe the specific details.

Regression model establishment for each energy end use

This project constructed 4 multi-regression models for each home energy end use, water heating, space heating, space cooling, and appliances, choosing the independent variables from RECS dataset and taking each end use category as dependent variable. Specifically, combined with associated independent variables such as housing unit features (e.g. number of rooms, square feet), householder's characteristics (e.g. income, race), and regional influences (e.g. heating degree days), ordinary least square (OLS) is employed to build the multi-regression model. Generally, each established model can be formulated as:

$$lnC_m = A_m + \sum_n (B_{mn} \times R_{mn}) \quad (1)$$

where C_m indicates the total annual energy consumption of category m, per household; A_m indicates the constant value for established regression model of end use category m; R_{mn} indicates the chosen variable n from RECS, accounting for end use category m; B_{mn} indicates the coefficient of variable R_{mn}.

The reason that those regression models were constructed as log-linear type, as shown in formula (1), is the relatively higher adjusted R-squared value[4], which means a better prediction of lower error (see Table 2). Moreover, some selected independent variables were also transformed to other forms, such as logarithm, to better account for the dependent variable (i.e., energy consumption for each end

[4] *It indicates the percentage of the values of dependent variable that is predicted by independent varibales.*

use). Later, the established models are used to estimate both regional and individual home energy consumption of each end use, by using census data and personalized input data, respectively.

Variable selection

To construct models that best account for the energy end uses of each category, it is necessary to select variables of high relevance to the associated end use category. Space heating, for example, is closely related to the household area because it is intuitive that with bigger house individual tends to face a higher electricity bill due to a greater demand for space heating. In particular, space heating, space cooling, water heating, and appliances end use regression model have 13, 11, 12, 14 variables, respectively (see appendix for abbreviation glossary and specific variables of each model). Summary of variable categories and selected variables are shown as follows:

- Householder: Householder's age, and household's gross annual income.
- Regional factor: heating/cooling degree days, location division.
- Housing unit: Number of rooms, total square foot, and construction year.

Space heating model specification

Space heating model regresses on multiple independent variables that are closely associated with space heating pattern (see appendix), such as heating degree days and total square foot. Dependent variable is energy consumption for space heating, in kilowatt hour (kWh), which is directly obtained from corresponding RECS variable. In particular, this model only involves energy sources of natural gas and electricity, due to the fact that they comprise 77% of total energy consumed for space heating (DOE 2010b). Though RECS does not contain electricity price, this approach estimate it through dividing the total space heating energy consumption (KWHSPH) by the associated energy expense, both of which are provided by RECS.

Space cooling model specification

Considering the similarity to space heating, space cooling model regresses on multiple independent variables that are similar to those of space heating model, such as cooling degree days and total square foot. Dependent variable is energy consumption for space cooling, in kWh, which is directly obtained from corresponding RECS variable. Though RECS does not contain electricity price, this approach estimate it through dividing the total space cooling energy consumption (KWHSPC) by the associated energy expense, both of which are provided by RECS.

Water heating model specification

Unlike space heating and cooling, variables from RECS are not directly related to water heating usage pattern. For example, RECS does not contain variable of number of occupants, and usually it is directly associated with water usage, which also influences energy consumption for water heating. But variables from RECS that are indirectly linked to water heating can be employed as implication to account for this end use. For instance, number of rooms is used as a substitution to explain the number of occupants. This water heating model is constructed in this variable selection criterion (see Appendix). Also, though RECS does not contain electricity price, this approach estimate it through dividing the total water heating energy consumption (KWHWTH) by the associated energy expense, both of which are provided by RECS.

Appliances model specification

As for appliances model, RECS does not specifically estimate appliances end use, instead RECS estimate energy end uses in refrigerator and other sources (excluding space heating, cooling, water heating, and refrigerator). This appliances model combines the estimated energy use of refrigerator and other sources provided by RECS, and thus creating a dependent variable for appliances end use. This accounts for an unusually high energy consumption of appliances in the prediction model. Though RECS does not contain electricity price, this approach estimate it through dividing the total aggregation consumption by the associated energy expense, both of which are provided by RECS. Other independent variables are directly from RECS and the selection standard is much similar to that of water heating model. Number of occupants, for example, is closely related to the stove usage frequency, and also number of rooms is employed as its substitution.

Energy end use estimates by established model

By personalized input

An excel-based model is also built to provide individual user an approach to get more accurate end-use estimates by inputting individual information for each variable. Specifically, each end use estimate is calculated through the following equation:

$$E_m = e^{A_m + \sum_n (B_{mn} \times I_{mn})} \quad (2)$$

where E_m indicates the estimated total annual energy consumption of category m, per household; A_m indicates the constant value for established regression model of end use category m; I_{mn} indicates the variable chosen variable n from RECS for each end use model, whose value comes from individual input; B_{mn} indicates the coefficient of variable I_{mn}.

By applying the formula (2) derived from established regression models, this excel will produce a pie chart as well as a column chart, explaining the share and absolute value of each end use, respectively (see Figure 3)

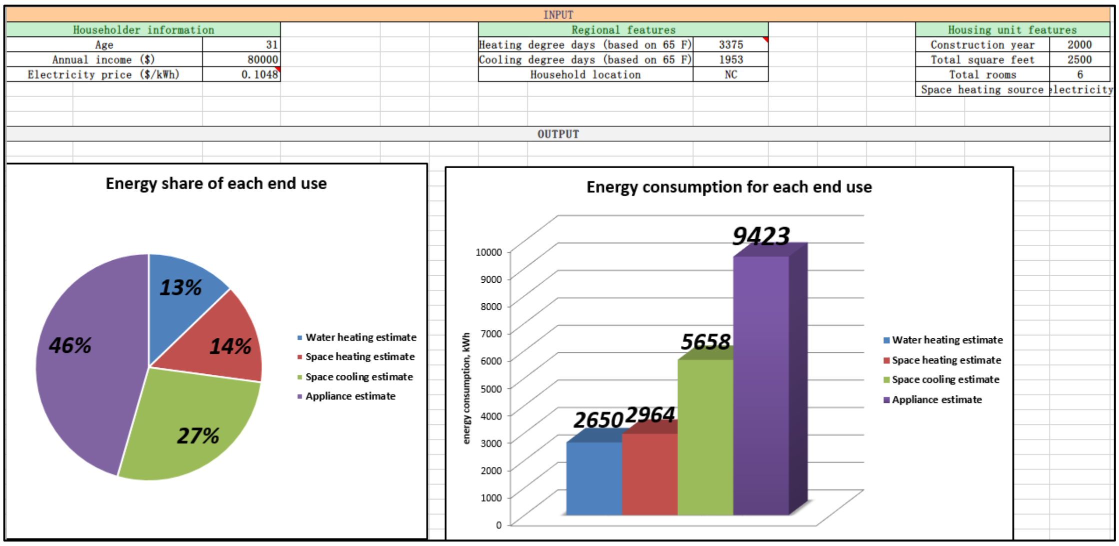

Figure 3: Example of estimation model for individual home energy end uses

By census data (Durham)

Combining formula (2) and variable data from U.S. Census and other sources[5], estimates are produced for Durham at a zip-code resolution. In terms of zip-code resolution estimates, each zip-code

[5] *Associated data were retrieved from American FactFinder of U.S Census website. Average price per square foot is sourced from http://www.trulia.com/.*

area will be associated with estimates of four end use categories, because each variable will be plugged in with value that are from census data of a specified zip-code. Specifically, total square foot is calculated through dividing housing unit value by average price for chosen zip-code area, because of U.S. Census' deficiency of providing total square foot. As for the value selection of other variables, if provided by U.S. Census, median value are selected to mitigate the outlier effect (e.g. HHAGE, MONEYPY), while if not, value of highest percentage will be selected as a representative (e.g. YEARMADE). Average retail price of electricity is employed (EIA 2013b); HDD65 and CDD65 are calculated as 5-year average[6].

By applying the values to the established model, end-use estimates for Durham at a zip-code resolution are calculated, as indicated in Figure 4.

Figure 4: End-use estimates for each zip code of Durham

Since U.S. Census provides data at different resolution-level, ranging from county to block, this approach can alternatively produce estimation database corresponding to the input data resolution.

[6] *Calculation was conducted through www.degreedays.net*

RESULTS

Overview

After finishing the model construction[7], another issue is how to use the model and established database (i.e., zip-code level database produced by these models) to reduce an individual's home energy consumption. This section begins with several energy efficiency upgrades targeting at different end uses, and a step-by-step tutorial example on how to make use of these models are also provided to illustrate the practical use of this project.

Home energy efficiency upgrades on each end use

As Department of Energy's (DOE) Buildings Energy Data Book indicates, for household using electricity as their primary energy source, heating and cooling accounts for 43% of energy consumption, while lighting and water heating consume 9.7% and 12.9%, respectively (DOE 2010c). Energy efficiency retrofits for different end uses could have distinctive effect due to the share of each energy end use. Hence, after individual estimates the energy consumption for each end use, it is crucial to apply efficiency improvements for certain end use categories where they can be most cost-effective, especially given a limited budget. For each energy conservation advice, the following aspects will be illustrated to better provide individual with recommendations that can best satisfy their needs:

- Technology overview
- Strengths of recommendation
- Drawbacks of recommendation

Space heating and cooling

Space heating and cooling consumes the largest portion of energy in U.S. buildings, with a share of 46.6% (DOE 2010a). Among a variety of factors affecting the space heating and cooling energy usage, some of them can hardly decrease due to uncontrollable reasons (e.g. climate factors), while some could be potentially reduced through proper methods.

[7] *In the Appendix section, each term in formula (1) is summarized in an associated table of each model to better interprete the model specifications. In addition, model details (STATA outputs) are also provided*

Insulation

Heat loss is primarily due to heat transfer, including conduction, convection, and radiation. For residential housing unit, passage of fluid such as air around the house can strengthen the convection effect, reducing/raising the indoor temperature in accordance to outside temperature. Insulation focuses on reducing any heat transfer between inside and outside of house by using insulation materials to seal areas of higher heat transfer ability, such as shafts, exterior doors and windows, and recognized spots of leakage. Resident can improve space heating/cooling efficiency by referring to different leakage potential of each source (Figure 5), as well as the insulation efficiencies of various materials (Table 1).

It is estimated that insulation and air sealing could contribute a 10% cut in energy cost, in individual homes (Energy Star n.d.). In addition to the saving potential, insulation could also be gradually accomplished without facing too much upfront cost (Jia et al. 2012). Because of the housing unit characteristics, each leakage spot can be sealed without necessarily affecting other spots. For example, individual can achieve an energy efficiency upgrade by solely sealing doors without other insulations. Easy as it seems to be, insulation work might need contactors to keep working hours or days to finish, presenting disturbance to residents.

Table 1: Insulation materials[8]

Insulation material	Material grouping
Polyurethane	Honeycombed
Extruded polystyrene	Honeycombed plastic
Cork	Vegetable−based
Linen wool	Vegetable−based
Rock wool	Mineral base
Glass wool	Mineral base

[8] *Listed materials are in decreasing order of insulation efficiency. Retrieved from International Energy Agency 2007*

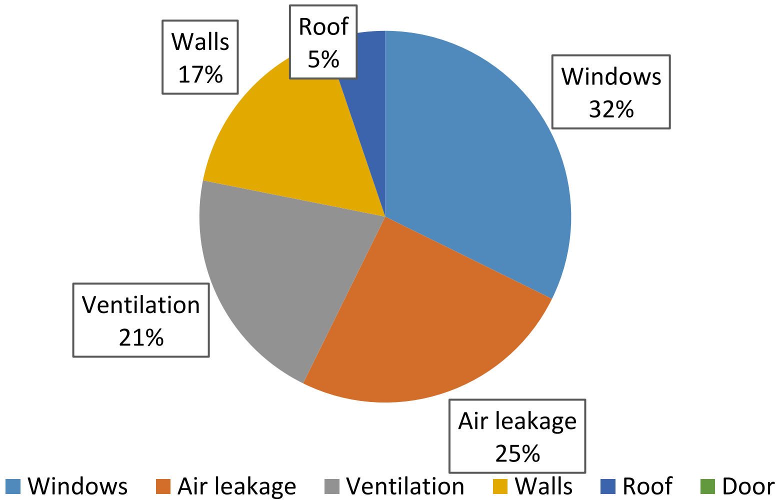

Figure 5: Heat loss sources in multi-unit residential buildings (Canada Mortgage and Housing Corporation 2007)

Heating, ventilation and air conditioning (HAVC) system

HAVC system has been playing an indispensable role in maintaining a comfortable indoor temperature, and it is also wildly employed in modern buildings, consuming a great share of total residential energy use. RECS indicates that around 87% of apartment buildings in the United States are installed with air conditioners (EIA 2011)

A primary way to save energy on space heating/cooling sector is switching to newer HAVC system, which usually has higher energy efficiency ratings. By applying newer model, it is estimated that buildings without any upgrade in HAVC could achieve a 30% efficiency improvements, and the payback period is also acceptable (Jia et al. 2012). In addition to upgrading HAVC system, insulation such as caulking the windows could also reduce the required energy to maintain a desired temperature. Though upgrading HAVC system might be a cost-effective option in terms of its acceptable payback period, it may also be hindered by a relatively high up-front cost, especially given a tight budget.

Water heating

For residential sector, water is typically heated by water heater using diverse energy sources. Common fuels include electricity, natural gas, propane, and heating oil. Efficiency of water heating could differ because of the specific characteristics of each fuel source. For instance, of the entire US residential stock of water heaters in 2005, the input-energy efficiency of electricity-powered heater and gas-fired heater are 88% and 56%, respectively (DOE 2010d). Many options could be employed to achieve energy improvement, such as tankless water heater (Jia et al. 2012), and this recommendation will mainly focus on the application of solar water heater.

Solar water heater

Among several types of solar water heater, the most common one contains a solar collector and one or multiple storage tanks (DOE 2013). Solar water heater systems have either active or passive design, and the detailed system classification is shown in Figure 6. In particular, indirect circulation system is best suited in colder climate, while passive system is preferred in warmer climates (Jia et al. 2012).

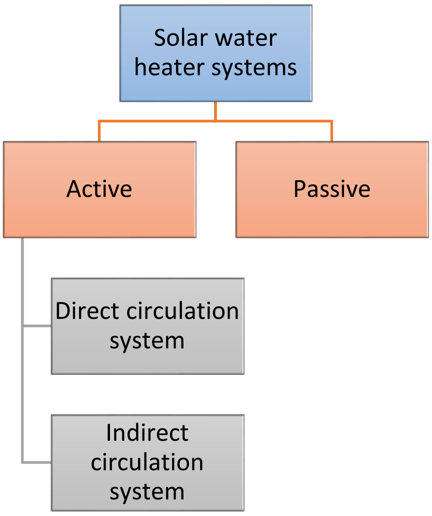

Figure 6: Solar water heater system classification[9] (Jia et al. 2012)

One great incentive for the solar heater system is the free and predictable energy source. But for some region of limited solar resource, this system cannot assure a continuous service, and thus a backup system is necessary. Additionally, a backup system could also provide capacity to meet hot water demand in peak usage times (Jia et al. 2012).

[9] *Direct circulation system uses solar colletor to circulate water as oppose to indircte circulation system that uses a heat-transfer fluid to transfer the heat from collector to water. Passive solar system take advantage of nature of water, the flow-tendency under heated condition.*

Example of how to model and reduce home energy consumption

As mentioned above, these models can be employed to estimate home energy consumption for both individual and the associated zip-code area. Therefore, the underlying idea of this project is to compare the individual estimates, given personal information input, with the zip-code average energy consumption of the individual's region.

Assume an individual and the associated house has the following characteristics: 1) 31-year old; 2) annual income is $80000; 3) zip-code of household location is 27705; 4) housing unit was constructed in 2000, with the square feet of 2500; 5) it has 6 rooms; 6) primary heating energy source is electricity. Inputting these information to the individual estimation model[10] and the end use estimation will be generated, as indicated in Figure 7.

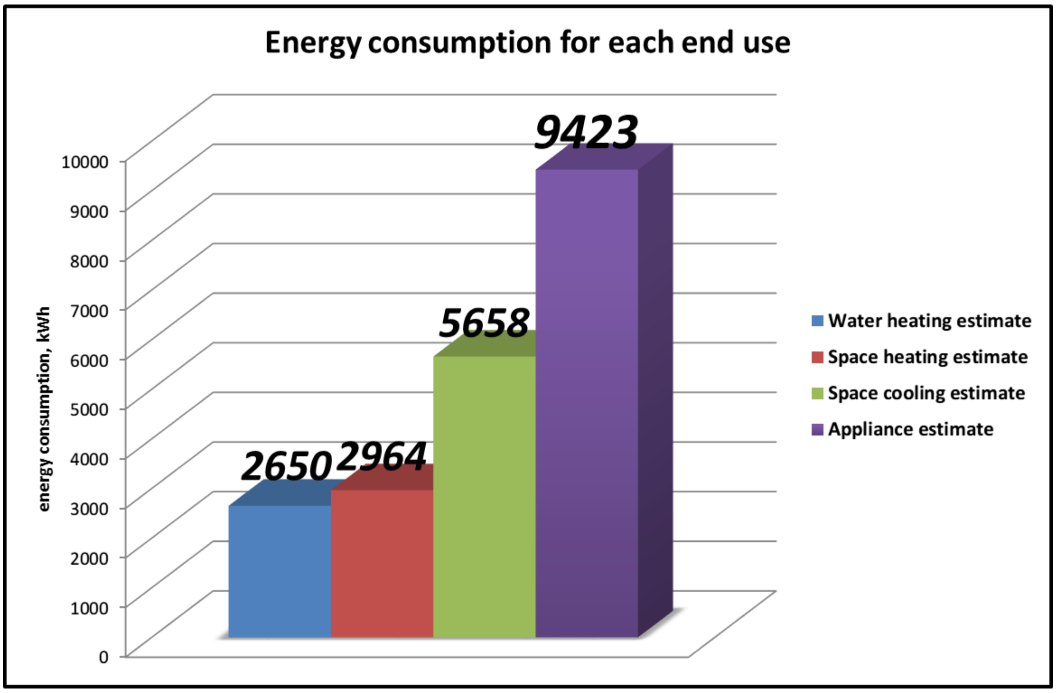

Figure 7: End use estimation by individual input

Meanwhile, the zip-code based database also displays the average end use details for zip-code 27705, as shown in Figure 8, which indicates this individual has a strangely high energy consumption in appliance and space cooling end use, suggesting that he/she may employ several insulation strategies (e.g., using polyurethane, material of the highest insulation efficiency according to Table 1, to

---

[10] *This model can be downloaded at: http://now4live.com/master-project/individual-estimate-V4.xlsx*

seal spots of potential leakage).

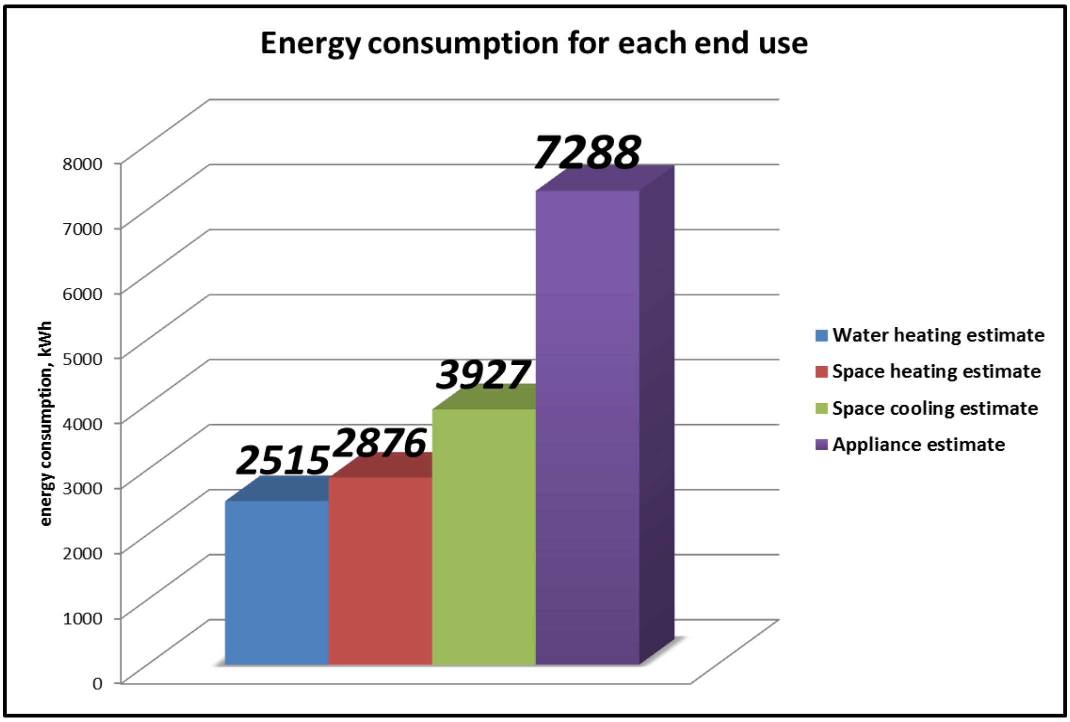

Figure 8: End use estimation of zip-code 27705

SUMMARY

By using the RECS as source dataset, this project constructed four end-use estimation models to estimate the home energy consumption by end uses. In addition, energy efficiency upgrades of each end use are also provided as the supplement of energy estimation. Overall, models has satisfactory predictability (i.e., adjusted R-squared value and P-value of each independent variable) and they can be further used as template to estimate energy end use consumption for both individual and certain zip-code areas.

Though the estimate result could be biased due to the limited input information, it is precious in showing individuals with the end uses of unusual energy consumption, compared to the averaged energy consumption in the same area. In all, in the absence of accurate estimates that needs various equipment and data input, this approach is valuable to give average users a sense of the magnitude of their energy consumption of each end use.

ACKNOWLEDGEMENT

The model methodology is adopted from a published article (Min et al. 2010) with updated source datasets, providing refined models of higher predictability. Dr. Peter Haff, my master project advisor, has been guiding the direction of my research as well as offering practical suggestions on how to better accomplish this project, with which I polished the original articles and combined the results with instructions that help average reader use the model to reduce the unnecessary home energy consumption. My client Mr. Rob Pinder also brought up this idea and offered me enormous assistances in each stage. Special thanks to Nicholas School of the Environment providing STATA software to run and build those models, and the feedbacks from both practice and final presentation further allowed me to better complete this project and revise this article.

APPENDIX I

Variable glossary

Abbreviation	Description
KWHWTH	Kilowatt hour (kWh) used for water heating per surveyed household
KWHSPH	kWh used for space heating per surveyed household
KWHCOL	kWh used for space cooling per surveyed household
KWHAPPL	kWh used for appliances per surveyed household
HDD65	Heating degree days in 2009, base temperature 65F
CDD65	Cooling degree days in 2009, base temperature 65F
YEARMADE	Year housing unit was built
YEARMADERAGE	Year range when housing unit was built[11]
MONEYPY	2009 gross household income[12], $
HHAGE	Age of householder
TOTSQFT	Total square footage (includes all attached garages, all basements, and finished/heated/cooled attics)
TOTROOMS	Total number of rooms in the housing unit
D1	New England Census Division (CT, MA, ME, NH, RI, VT) (YES=1, NO=0)
D2	Middle Atlantic Census Division (NJ, NY, PA) (YES=1, NO=0)
D3	East North Central Census Division (IL, IN, MI, OH, WI) (YES=1, NO=0)
D4	West North Central Census Division (IA, KS, MN, MO, ND, NE, SD) (YES=1, NO=0)
D5	South Atlantic Census Division (DC, DE, FL, GA, MD, NC, SC, VA, WV) (YES=1, NO=0)
D6	East South Central Census Division (AL, KY, MS, TN) (YES=1, NO=0)
D7	West South Central Census Division (AR, LA, OK, TX) (YES=1, NO=0)

[11] *1=Before 1950; 2=1950 to 1959; 3=1960 to 1969; 4=1970 to 1979; 5=1980 to 1989; 6=1990 to 1999; 7=2000 to 2004; 8=2005 to 2009*

[12] *MONEYPY contains 24 segments, and each segment has a range variance of $2500, $5000, or $10000.*

D8	Mountain North Sub-Division (CO, ID, MT, UT, WY) (YES=1, NO=0)
D9	Mountain South Sub-Division (AZ, NM, NV) (YES=1, NO=0)
D10	Pacific Census Division (AK, CA, HI, OR, WA) (YES=1, NO=0)
EPSPH	Electricity price for space heating, $/ kWh
EPSPC	Electricity price for space cooling, $/ kWh
EPWH	Electricity price for water heating, $/ kWh
EPAPPL	Electricity price for appliances, $/ kWh
SPHFN	Using natural gas as space heating energy (YES=1, NO=0)
SPHFE	Using electricity as space heating energy (YES=1, NO=0)
D3SPHTN	Space heating energy consumption (in thousand BTU) of housing unit using natural gas, in East North Central Division (D3).
D3SPHTE	Space heating energy consumption (in thousand BTU) of housing unit using electricity, in East North Central Division (D3).

Table 2: Summary of water heating estimate model[13]

Variable abbreviation[14]	Variable type	Coefficient	P-value
logKWHWTH	dependent variable	N/A	N/A
YEARMADE	independent variable	.0038631	0.000
HDD65	independent variable	.0000105	0.003
YEARMADERANGE	independent variable	-.0598652	0.000
TOTSQFT	independent variable	-.0000506	0.000
logTOTSQFT	independent variable	.109418	0.000
TOTROOMS	independent variable	.0790608	0.000
D1	independent variable	.2201888	0.000
D2	independent variable	.2362719	0.000
D7	independent variable	.1157279	0.000

---

[13] *Summary table exclude those variable of relatively low P –value (<5%), because a P-value of 5% or less is the generally accpetable point where this independent varible has some effect, at a 95% probability.*

[14] *See Appendix for glossary of variable abbreviation*

D8	independent variable	-.1703522	0.002
HHAGE	independent variable	-.0063405	0.000
EPWTH	independent variable	-5.789036	0.000

Table 3: Summary of space heating estimate model

Variable abbreviation	Variable type	Coefficient	P-value
logKWHSPH	dependent variable	N/A	N/A
HDD65	independent variable	-.0002535	0.000
sqrtHDD65	independent variable	.0507469	0.000
TOTSQFT	independent variable	-.0000528	0.000
logTOTSQFT	independent variable	.1073632	0.000
TOTROOMS	independent variable	.0267653	0.000
D2	independent variable	.0859133	0.002
D5	independent variable	.0972181	0.000
D6	independent variable	.0950842	0.000
D7	independent variable	.2154071	0.000
D8	independent variable	-.2535969	0.000
EPSPH	independent variable	-4.502746	0.000
SPHFN	independent variable	-.1415351	0.000
SPHFE	independent variable	1.549331	0.000

Table 4: Summary of space cooling estimate model

Variable abbreviation	Variable type	Coefficient	P-value
logKWHCOL	dependent variable	N/A	N/A
CDD65	independent variable	-.0013059	0.000
sqrtCDD65	independent variable	.175689	0.000
MONEYPY	independent variable	.0249616	0.000

logMONEYPY	independent variable	-.0884807	0.001
TOTSQFT	independent variable	.0000699	0.000
logTOTSQFT	independent variable	.4207122	0.000
TOTROOMS	independent variable	.0506229	0.000
D1	independent variable	-.2436908	0.000
D8	independent variable	-.2247187	0.000
D10	independent variable	-.4675279	0.000
EPCOL	independent variable	-4.296843	0.000

Table 5: Summary of appliances estimate model

Variable abbreviation	Variable type	Coefficient	P-value
logKWHAPPL	dependent variable	N/A	N/A
MONEYPY	independent variable	.0100529	0.000
TOTSQFT	independent variable	-.0001019	0.000
logTOTSQFT	independent variable	.4880063	0.000
TOTROOMS	independent variable	.0890411	0.000
D1	independent variable	.1186781	0.000
D2	independent variable	.1721214	0.000
D3	independent variable	.0968306	0.000
D4	independent variable	.0711206	0.000
D5	independent variable	.2061018	0.000
D6	independent variable	.2499598	0.000
D7	independent variable	.2483739	0.000
D9	independent variable	.1154287	0.000
HHAGE	independent variable	-.002971	0.000
EPAPPL	independent variable	-4.019902	0.000

Regression model specifications for water heating by STATA

Source	SS	df	MS
Model	432.849064	12	36.0707554
Residual	1096.42849	4800	.228422601
Total	1529.27755	4812	.317804977

Number of obs	=	4813
F(12, 4800)	=	157.91
Prob > F	=	0.0000
R-squared	=	0.2830
Adj R-squared	=	0.2812
Root MSE	=	.47794

logKWHWTH	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
HDD65	.0000105	3.56e-06	2.96	0.003	3.55e-06	.0000175
YEARMADE	.0038631	.001087	3.55	0.000	.0017322	.005994
YEARMADERANGE	-.0598652	.0122776	-4.88	0.000	-.0839349	-.0357954
TOTSQFT	-.0000506	.0000134	-3.77	0.000	-.0000769	-.0000243
logTOTSQFT	.109418	.0306475	3.57	0.000	.0493349	.169501
TOTROOMS	.0790608	.0052652	15.02	0.000	.0687387	.089383
D1	.2201888	.0369073	5.97	0.000	.1478335	.292544
D2	.2362719	.0323753	7.30	0.000	.1728015	.2997422
D7	.1157279	.0220483	5.25	0.000	.0725031	.1589526
D8	-.1703522	.0548358	-3.11	0.002	-.2778554	-.0628489
HHAGE	-.0063405	.0004123	-15.38	0.000	-.0071487	-.0055323
EPWTH	-5.789036	.2034768	-28.45	0.000	-6.187944	-5.390129
_cons	.1391229	2.102181	0.07	0.947	-3.982115	4.260361

Regression model specifications for space heating by STATA

Source	SS	df	MS
Model	3987.92883	13	306.763756
Residual	1358.55662	6140	.221263293
Total	5346.48545	6153	.868923362

Number of obs	=	6154
F(13, 6140)	=	1386.42
Prob > F	=	0.0000
R-squared	=	0.7459
Adj R-squared	=	0.7454
Root MSE	=	.47039

logKWHSPH	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
HDD65	-.0002535	.0000145	-17.42	0.000	-.000282	-.0002249
sqrtHDD65	.0507469	.0016045	31.63	0.000	.0476015	.0538922
TOTSQFT	-.0000528	.0000109	-4.83	0.000	-.0000742	-.0000313
logTOTSQFT	.1073632	.0259533	4.14	0.000	.0564857	.1582408
TOTROOMS	.0267653	.0044671	5.99	0.000	.0180083	.0355223
D2	.0859133	.0276632	3.11	0.002	.0316838	.1401427
D5	.0972181	.0172164	5.65	0.000	.0634678	.1309683
D6	.0950842	.0257009	3.70	0.000	.0447014	.145467
D7	.2154071	.0218456	9.86	0.000	.1725822	.2582321
D8	-.2535969	.0402019	-6.31	0.000	-.3324067	-.1747872
EPSPH	-4.502746	.1755247	-25.65	0.000	-4.846836	-4.158656
SPHFN	-.1415351	.0231963	-6.10	0.000	-.187008	-.0960621
SPHFE	1.549331	.0227941	67.97	0.000	1.504647	1.594015
_cons	3.858467	.1789192	21.57	0.000	3.507723	4.209211

Regression model specifications for space cooling by STATA

Source	SS	df	MS
Model	13093.3913	11	1190.3083
Residual	4673.77265	9928	.470766786
Total	17767.1639	9939	1.78762088

Number of obs	= 9940
F(11, 9928)	= 2528.45
Prob > F	= 0.0000
R-squared	= 0.7369
Adj R-squared	= 0.7367
Root MSE	= .68612

logKWHCOL	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
CDD65	-.0013059	.0000349	-37.43	0.000	-.0013743	-.0012375
sqrtCDD65	.175689	.0029015	60.55	0.000	.1700016	.1813765
MONEYPY	.0249616	.0027224	9.17	0.000	.0196251	.0302981
logMONEYPY	-.0884807	.0265203	-3.34	0.001	-.1404658	-.0364956
TOTSQFT	.0000699	.0000119	5.89	0.000	.0000466	.0000932
logTOTSQFT	.4207122	.0298887	14.08	0.000	.3621243	.4793001
TOTROOMS	.0506229	.0050777	9.97	0.000	.0406695	.0605763
D1	-.2436908	.0306937	-7.94	0.000	-.3038568	-.1835249
D8	-.2247187	.0436898	-5.14	0.000	-.3103596	-.1390779
D10	-.4675279	.0224177	-20.86	0.000	-.5114711	-.4235846
EPCOL	-4.296843	.1648871	-26.06	0.000	-4.620055	-3.973631
_cons	-.6616925	.2023272	-3.27	0.001	-1.058295	-.2650902

Regression model specifications for appliances by STATA

Source	SS	df	MS
Model	2476.04989	14	176.860706
Residual	3029.06963	12068	.251000135
Total	5505.11952	12082	.455646377

Number of obs	= 12083
F(14, 12068)	= 704.62
Prob > F	= 0.0000
R-squared	= 0.4498
Adj R-squared	= 0.4491
Root MSE	= .501

logKWHAPPL	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
MONEYPY	.0100529	.0007933	12.67	0.000	.0084979	.0116079
TOTSQFT	-.0001019	7.81e-06	-13.05	0.000	-.0001172	-.0000866
logTOTSQFT	.4880063	.0192535	25.35	0.000	.4502664	.5257463
TOTROOMS	.0890411	.0033562	26.53	0.000	.0824625	.0956197
D1	.1186781	.020227	5.87	0.000	.0790299	.1583263
D2	.1721214	.0181132	9.50	0.000	.1366166	.2076262
D3	.0968306	.0180655	5.36	0.000	.0614193	.1322419
D4	.0711206	.016263	4.37	0.000	.0392425	.1029987
D5	.2061018	.0146783	14.04	0.000	.1773299	.2348737
D6	.2499598	.0228373	10.95	0.000	.2051949	.2947246
D7	.2483739	.0175015	14.19	0.000	.2140681	.2826797
D9	.1154287	.0280308	4.12	0.000	.0604837	.1703736
HHAGE	-.002971	.0002792	-10.64	0.000	-.0035183	-.0024237
EPAPPL	-4.019902	.1201672	-33.45	0.000	-4.255449	-3.784355
_cons	5.169526	.1230322	42.02	0.000	4.928364	5.410689

APPENDIX II

Log-linear model versus linear model

In the analysis above, log-linear statistical model[15] is chosen as the template of estimate model, strengthening those models in the following aspects:

1. Adjusted R-squared value: Log-linear models have higher adjusted R-squared value (as indicated in Table 2), given same independent variables in each estimate model, meaning that higher percentage values of dependent variable are accounted for, compared to linear model.
2. Normality testing: Log-linear models show a less strong tendency of deviating the straight line than that of linear models (as indicated in Figure 6), meaning that log-transformed values of dependent variable are more normally distributed.

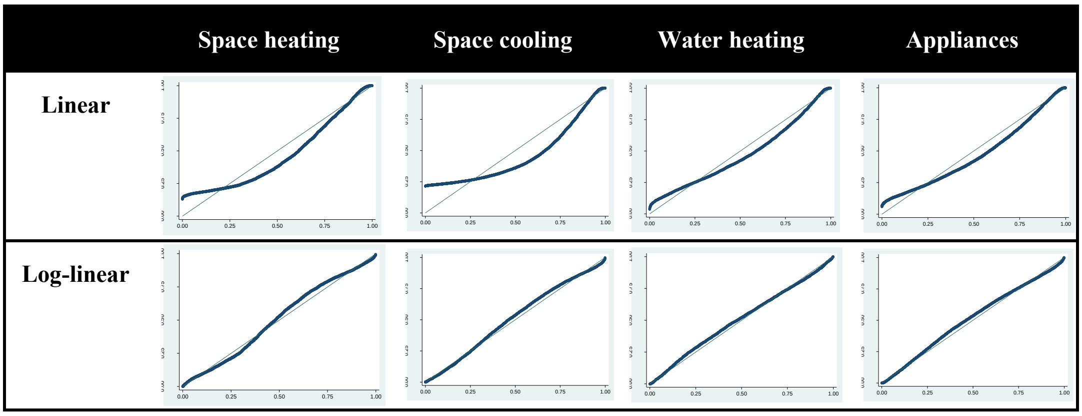

Figure 9: Standardized normal probability plot of each end use variable in RECS[16]

Table 6: Comparison of Adjusted R-squared value of linear and non-linear model

Model	Space heating	Space cooling	Water heating	Appliances
Log-linear	0.7458	0.7382	0.2854	0.4493
Linear	0.5916	0.5632	0.2113	0.3566

[15] *Transforming the dependent variable to logriathm form rather than simply using the original value as denependent varible (i.e. linear model).*

[16] *In each chart, horizontal axis indicates standrized residual, while vertical axis indicates normal % probablilty*

Potential errors in RECS survey data

The accuracy of end use models is greatly dependent on the RECS dataset, as the multi-regression was based on the values of a variety of RECS variables. Specifically, dependent variables in each model (i.e. estimate of each end use) are estimated by EIA, as mentioned in previous section. According to the published data (DOE n.d.), the efficiency difference between using natural gas and electricity falls around 60% to 70%, but the RECS estimates dramatically exceed this range, as indicated in Table 3, presenting a potential error in RECS estimation methodology.

Table 7: Comparison of space heating energy usage between natural gas and electricity[17]

	Mean	Standard error	95% confident interval	
D3PHTN	2658.42	130.90	2400.51	2916.33
D3PHTE	13863.90	576.78	12725.28	15002.53

[17] The result is produced through calculating the mean of space heating energy consumption (in thousand BTU) between natural gas and electricity, in the same survey region (D3).

REFERENCE

Canada Mortgage and Housing Corporation. 2007. *Air Leakage Control Manual for Existing Multi-Unit Residential Buildings*. Retrieved (http://www.cmhc-schl.gc.ca/odpub/pdf/65847.pdf).

DOE. 2010a. "1.1.4 2010 U.S. Buildings Energy End-Use Splits, by Fuel Type (Quadrillion Btu)." in *Buildings Energy Data Book 2010*. Retrieved (http://buildingsdatabook.eren.doe.gov/TableView.aspx?table=1.1.4).

DOE. 2010b. "2.1.1 Residential Primary Energy Consumption, by Year and Fuel Type (Quadrillion Btu and Percent of Total)." in *Buildings Energy Data Book 2010*. DOE.

DOE. 2010c. "2.1.5 2010 Residential Energy End-Use Splits, by Fuel Type (Quadrillion Btu)." in *Buildings Energy Data Book 2010*.

DOE. 2010d. "5.4.5 Water Heater Efficiencies." Pp. 5–17 in *Buildings Energy Data Book 2010*. Retrieved (http://buildingsdatabook.eren.doe.gov/docs\DataBooks\2010_BEDB.pdf).

DOE. 2013. "Solar Water Heater Basics." Retrieved (http://energy.gov/eere/energybasics/articles/solar-water-heater-basics).

DOE. n.d. "Electric Resistance Heating." Retrieved (http://energy.gov/energysaver/articles/electric-resistance-heating).

EIA. 2009. "How Does EIA Estimate Energy Consumption and End Uses in U.S. Homes?" Retrieved (http://www.eia.gov/consumption/residential/reports/2009/methodology-end-use.cfm).

EIA. 2011. "RESIDENTIAL ENERGY CONSUMPTION SURVEY (RECS) Analysis and Projections." Retrieved (http://www.eia.gov/consumption/residential/reports.cfm).

EIA. 2013a. "Annual Energy Outlook 2013."

EIA. 2013b. *Electric Power Monthly with Data for December 2013*. Retrieved (http://www.eia.gov/electricity/monthly/pdf/epm.pdf).

EIA. 2014. *AEO2014 Early Release Overview*. Retrieved (http://www.eia.gov/forecasts/aeo/er/pdf/tbla2.pdf).

EIA. n.d. "About RECS." Retrieved (http://www.eia.gov/consumption/residential/about.cfm).

Energy Star. n.d. "Methodology for Estimated Energy Savings from Cost-Effective Air Sealing and Insulating." Retrieved (http://www.energystar.gov/index.cfm?c=home_sealing.hm_improvement_methodology).

Jia, Aimee, Thomas Opp, David Smedick, Allison Smykal, and Jason Symonds. 2012. *NYC CO-OP AND CONDOMINIUM BOARD GUIDE TO ENERGY EFFICIENCY UPGRADES IN BUILDINGS*. Retrieved (http://dukespace.lib.duke.edu/dspace/handle/10161/5317).

Min, Jihoon, Zeke Hausfather, and Qi Feng Lin. 2010. "A High-Resolution Statistical Model of Residential Energy End Use Characteristics for the United States." *Journal of Industrial Ecology* 14(5):791–807. Retrieved January 31, 2014 (http://doi.wiley.com/10.1111/j.1530-9290.2010.00279.x).

Saidur, R., H. H. Masjuki, and M. Y. Jamaluddin. 2007. "An Application of Energy and Exergy Analysis in Residential Sector of Malaysia." *Energy Policy* 35(2):1050–63. Retrieved January 23, 2014 (http://linkinghub.elsevier.com/retrieve/pii/S0301421506000930).

Seryak, John, and Kelly Kissock. 2003. "OCCUPANCY AND BEHAVIORAL AFFECTS ON RESIDENTIAL ENERGY USE." Pp. 717–22 in *In Proceedings of the Solar Conference*. AMERICAN SOLAR ENERGY SOCIETY; AMERICAN INSTITUTE OF ARCHITECTS.

Swan, Lukas G., and V. Ismet Ugursal. 2009. "Modeling of End-Use Energy Consumption in the Residential Sector: A Review of Modeling Techniques." *Renewable and Sustainable Energy Reviews* 13(8):1819–35. Retrieved January 20, 2014 (http://linkinghub.elsevier.com/retrieve/pii/S1364032108001949).