Energy & Buildings 240 (2021) 110934

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enb

An examination of factors affecting residential energy consumption using a multiple discrete continuous approach

Naveen Chandra Iraganaboina *, Naveen Eluru

Department of Civil, Environmental & Construction Engineering, University of Central Florida

ARTICLE INFO

Article history:
Received 17 October 2020
Revised 12 March 2021
Accepted 14 March 2021
Available online 19 March 2021

Keywords:
Residential energy
Energy mix
MDCEV
Residential energy consumption survey

ABSTRACT

Residential energy use has become an important source of global energy demand growth and carbon emissions growth. Residential building energy usage accounts for about 22% of the total energy use in the United States. In the current study, we address residential energy usage by addressing two decisions: (1) source of energy (such as electric and natural gas) and (2) consumption by energy source for various purposes. A Multiple Discrete Continuous Extreme Value model that allows us to analyze the source and consumption decisions in an integrated framework is developed. The model is estimated using data drawn from the 2015 Residential Energy Consumption Survey that provides energy use details for electricity, natural gas, fuel oil and Liquefied Petroleum Gas for residential units across United States. An exhaustive set of independent variables including location characteristics, household characteristics, housing characteristics, appliance use and climatic characteristics were employed in the model estimation. The model estimation results are augmented with a comprehensive policy analysis to illustrate how various independent variables affect energy use by source. A comparison of energy use between urban and rural regions, by varying household size and housing unit size are examined.

1. Introduction

United States of America uses around 17% of the world's annual energy usage with just 4.3% of world's population [1,2]. Energy usage and delivery patterns of households are undergoing a significant transformation. On the demand side this is being facilitated by advances in transportation technology (such as electric vehicles, connected autonomous vehicles, V2X communications), building science (smart and resilient buildings), computing (artificial intelligence, machine learning and mobile applications), and changing landscape of consumer choices (such as in-home activity participation with smart devices). On the delivery side, this has been spurred by energy mix (such as fossil fuels, coal, hydro, solar and wind energy), smart distribution systems, micro-grids, and complex economic and legal structures. Not only do these changes affecting demand and supply side have direct impacts but they also interact with each other in complex ways causing indirect and induced impacts. Therefore, there is a *need for modeling tools that allow for a holistic understanding of the energy demand–supply* to assist public agencies, utility companies and other stakeholders in making informed decisions for the future of energy infrastructure and delivery services. Towards building these holistic frameworks, the current research effort focuses on an important component of energy usage – residential building energy usage - that accounts for about 22% of the total energy use in the United States [3]. Residential energy use has become an important source of global energy demand growth and carbon emissions growth [4,5]. In our analysis, we focus on developing a residential energy use prediction framework for United States with a nationally representative household level energy usage data. As the share of electric vehicles increases within our transportation infrastructure, the spatio-temporal nature of current electricity demand is likely to alter with increased household electricity use for vehicle charging. To develop a future estimate of urban demand with electric vehicles, a model system of current use serves as a baseline estimate.

Understanding residential energy usage includes addressing two decisions: (1) source of energy (such as electric and natural gas) and (2) usage by energy source for various purposes. Across various urban regions in the United States, several residences have potentially multiple energy source options. For instance, in some regions, electricity is used for lighting, air conditioning and refrigeration while natural gas is used for cooking and heating. In this case, the household consumes energy from two sources and energy usage exists by source. Traditionally in energy literature, analyzing such data involved estimating separate models by energy source

* Corresponding author.
E-mail addresses: naveen.chandra@knights.ucf.edu (N.C. Iraganaboina), naveen.eluru@ucf.edu (N. Eluru).

https://doi.org/10.1016/j.enbuild.2021.110934
0378-7788/

Abbreviations

AC	Air Conditioning
ANN	Artificial Neural Network
BTU	British Thermal Unit
CDD	Cooling Degree Days
CER	Commission for Energy Regulation's
EIA	Energy Information Administration
FIES	Family Income and Expenditure Survey
HDD	Heating Degree Days
HH	Household
IRT	Item Response Theory
LPG	Liquid Petroleum Gas
MDC	Multiple Discrete Continuous
MDCEV	Multiple Discrete Continuous Extreme Value
MITYC	Ministry of Industry, Tourism and Trade
NVM	Dutch Realtor Association
PHEBUS	Housing Performances, Equipment, Needs, and Usages of Energy.
RAM-MLF	Resource Allocation Model based on Multi Linear Function
RECS	Residential Energy Consumption Survey
STSM	Structural Time Series Model
SVM	Support Vector Machine
TV	Television
V2X	Vehicle to Other
US	United States

that take the form of a linear regression (or its variants). While this approach is relatively straightforward and provides useful insights, it inherently ignores the interaction between the energy sources and usage patterns by energy source. The overall decision process can be more efficiently examined using a recently developed econometric model structure that allows for selecting multiple alternatives (and their associated usages). The framework labelled as Multiple Discrete Continuous (MDC) systems allow analysis of households selecting multiple options within a behaviorally elegant framework random utility framework. Several model structures have been proposed under the MDC realm. Of these the Multiple Discrete Continuous Extreme Value (MDCEV) model with its close form structure is the most adopted framework [6–8].

The MDCEV model system is estimated using data drawn from the 2015 Residential Energy Consumption Survey (RECS) that provides energy use details for electricity, natural gas, fuel oil and Liquefied Petroleum Gas (LPG) for residential units across United States. An exhaustive set of independent variables including location characteristics (census region, type of location), household characteristics (such as HH income, race, household size, education), housing characteristics (such as year of construction, housing type, type of unit, square footage, and number of stories), appliance use (such as appliances used in the housing unit) and climatic characteristics (such as heating degree days and cooling degree days) were employed in the model estimation. The model estimation results are augmented with a comprehensive policy analysis to illustrate how various independent variables affect energy use by source.

2. Earlier literature on residential energy

As expected, residential energy use is a well researched field (see Amasyali and El-Gohary [9] and Wei et al [10] for a detailed review). Earlier research in residential energy use focused on choice of energy source [5], clustering of energy profiles [11–13], effect of policy decisions on energy use [14], behavioral triggers for energy saving and energy use [14–20] across different segments of population. It is important to note that some of these studies focused on a subset of energy use choices (such as heating, lighting, cooling, and cooking). The research methodologies employed for analysis include classical regression models, discrete continuous models, Artificial Neural Networks (ANN) [21–25], Support Vector Machine (SVM) [22,25], Structural Time Series Model (STSM) [26–28], and Genetic Algorithms [29,30]. Given the focus of the current research effort in developing an advanced econometric model for energy usage analysis, we restrict our review of residential energy literature employing econometric modeling approaches. A summary of the relevant literature reviewed is presented in Table 1 with information on study region, data used for analysis, energy sources considered, dimensions considered (choice of energy source and/or usage by energy source), independent variables considered, and modeling approach employed.

From the Table, several observations can be made. *First*, two different types of data were used to understand the energy use behavior in residential dwellings. The first type of data includes retrospective survey data compiled by government institutions or researchers. The second type of data are obtained using monitoring devices to record the household energy use. *Second*, the geographical extent of the research covers various countries including Netherlands, Norway, China, United States, Ireland, India, Portugal, France, Spain, and Australia. *Third*, the independent variables considered include the following categories: location characteristics, household characteristics, housing characteristics, appliance use in the housing unit, and climatic characteristics. From the literature, it is evident that regional characteristics reveal the inherent differences in energy mix. The size and configuration of housing unit and household influence the energy use of energy alternatives. Climatic variables play a crucial role in usage of energy sources to meet heating and cooling needs. With the increase in the appliances used by the household, the energy use of the housing unit increases or decreases intuitively. *Fourth*, methodologies considered in these studies vary based on the dimension of interest–choice of energy source and/or usage by energy source. In the studies that involve selection of energy alternatives, categorical modeling approaches such as multinomial regression models are used. In studies focused on usage by energy alternative, models employed include linear regression models, lasso regression, log-linear and log–log linear regression models. In the studies that examine both dimensions, methodologies considered include MDCEV, resource allocation models and discrete–continuous econometric models were used. *Finally*, while a majority of the earlier research has focused on electricity usage, other energy sources such as natural gas, fuel oil and liquid petroleum gas have also been studied.

It is evident from the literature review that substantial research has been conducted to examine the relationship between various independent variables and usage of various energy sources. However, several important aspects related to energy source selection and usage are not fully understood. The proposed study contributes to the literature on residential energy along the following directions. The current study develops a unified framework for analyzing energy source selection and usage to address the interconnected nature of these decisions. Earlier research mainly employed model frameworks that are reliant on linear regression approaches. These approaches inherently are not suited to capture

Table 1
Literature review matrix.

Reference	Study region	Data	Energy sources	Dimension of Interest	Independent variables	Modelling approach
Anderson et al., 2017 [16]	Ireland	Irish Commission for Energy Regulation's (CER) Smart Metering Electricity Customer Behaviour Trials (CBTs)	Electricity	Usage	Household (HH) characteristics	Mixed effects linear regression model
Bedir et al., 2013 [17]	Netherlands	Survey data	Electricity	Usage	HH characteristics, housing characteristics, appliance use, economic characteristics	Linear regression
Belaid and Garcia, 2016 [14]	France	French PHEBUS dataset	Electricity	Usage	Location characteristics, HH characteristics, housing characteristics, climatic characteristics	Item Response Theory (IRT), Linear regression
Blázquez et al., 2013 [18]	Spain	Ministry of Industry, Tourism and Trade (MITYC)	Electricity	Usage	HH characteristics, climatic characteristics, population, price of electricity	Log-Log Linear regression
Brounen et al., 2012 [20]	Netherlands	Dutch Realtor Association (NVM) Data	Electricity, Natural Gas	Usage	HH characteristics, housing characteristics, climatic characteristics	Linear regression
Jones and Lomas, 2015 [31]	UK	Survey conducted in Leicester (UK)	Electricity	Usage	Appliance use	Odds ratio method
Kavousian et al., 2013 [32]	Silicon Valley (California)	Survey data	Electricity	Usage	HH characteristics, housing characteristics, climatic characteristics	Linear regression
Huebner et al., 2016 [33]	UK	Survey data	Electricity	Usage	HH characteristics, housing characteristics, appliances use	Lasso regression
Hori et al., 2013 [34]	Five different Asian cities	Survey data	Electricity	Usage	HH characteristics, energy use consciousness, environmental behaviour, social interaction	Linear regression
Huang, 2015 [35]	Taiwan	FIES (Family Income and Expenditure Survey)	Electricity	Usage	Location characteristics, HH characteristics, housing characteristics	Quantile regression
Pinjari and Bhat, 2011 [36]	USA	RECS	Electricity, Natural Gas, Fuel oil, LPG	Choice and Usage	Location characteristics, HH characteristics, housing characteristics, climatic characteristics	MDCEV
Wiesmann et al., 2011 [37]	Portugal	Instituto Nacionalde Estatistica (INE) database	Electricity	Usage	Location characteristics, HH characteristics, housing characteristics, climatic characteristics	Log-Linear model
Yu and Zhang, 2015 [38]	Beijing, China	Survey	Electricity (End use)	Choice and Usage	HH characteristics	MDCEV and Resource allocation model based on multi linear function (RAM-MLF)
Dale et al., 2009 [39]	USA	RECS	Electricity, Natural gas	Usage	HH characteristics, climatic variables characteristics, price of energy source	Log-Log Linear regression
Dubin and McFadden, 1984 [40]	USA	Survey by Washington Center for Metropolitan Studies	Electricity and Natural Gas (For heating)	Choice and Usage	Price and availability of energy source	Joint Multinomial Logit and Linear regression
Filippini and Pachauri, 2004 [41]	India	Survey data	Electricity	Usage	HH characteristics, price	Log-Linear regression
Mansur et al., 2008 [42]	USA	RECS	Electricity, Natural gas, Fuel oil and LPG	Choice and Usage	HH characteristics, climatic characteristics, price of energy source	Joint Multinomial Logit and Linear regression
Narayan and Smyth, 2005 [43]	Australia	International Energy Agency data	Electricity	Usage	HH characteristics, climatic characteristics, price	Log-Log demand model
Nesbakken, 2001 [44]	Norway	Norwegian micro data	Electricity, oil and wood (for heating)	Choice and Usage	Housing characteristics, climatic characteristics, price of energy source	Discrete – continuous choice model
Vaage, 2000 [45]	Norway	Norway energy survey data	Heating appliances	Choice and Usage	HH characteristics, appliances use, price of energy source	Multinomial Logit Model, Linear regression
Sailor and Muñoz, 1997 [46]	USA	RECS	Electricity and natural gas	Usage	Climatic characteristics	Linear regression
Harold et al., 2015 [47]	Ireland	Commission for Energy Regulation (CER) data	Natural gas	Usage	HH characteristics, housing characteristics, climatic characteristics	Linear regression

interactions across energy source selection and usage dimensions. Towards accommodating these interactions, we adopt the MDCEV model that has been widely applied in multiple discrete continuous decision contexts. Further, we also build on existing literature by considering an exhaustive set of independent variables from location characteristics, household characteristics, housing

Table 2
Description of Energy Consumed by Energy Alternatives.

Energy alternative	Availability	Selection	Mean usage (when energy source is selected) (in 10^6 BTU)	Mean usage (when energy source is selected) (in million Joules)
Electricity	100%	100%	37.73	39.81
Natural Gas	69.3%	58.5%	57.40	60.56
Fuel Oil	Not available	5.2%	70.50	74.38
Propane (or LPG)	Not available	10%	32.45	34.23

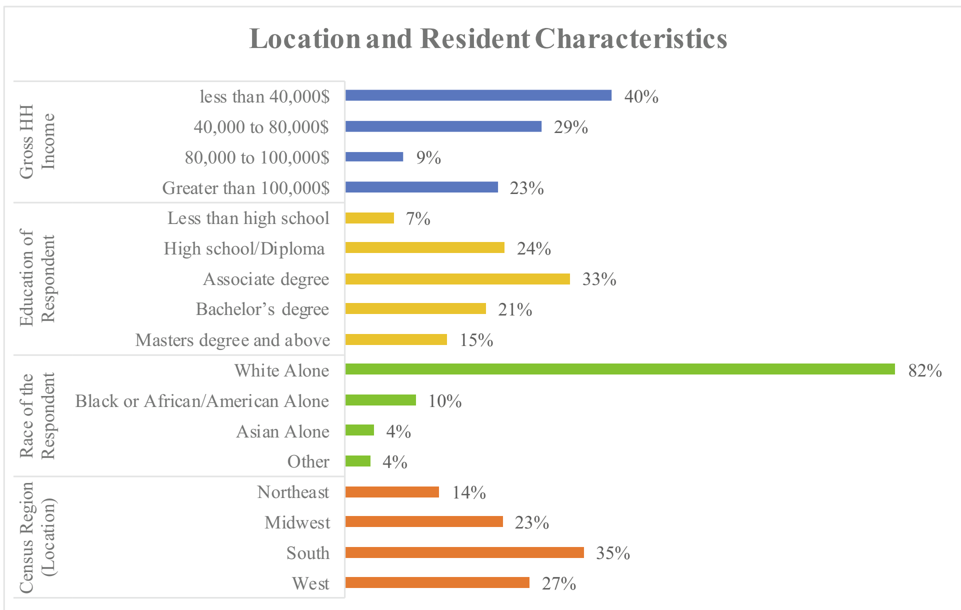

Fig. 1. Descriptives of respondent characteristics and location of residential dwelling.

characteristics, appliance use and climatic conditions, in examining the residential energy decision framework. The proposed research uses Residential Energy Consumption Survey (RECS) 2015 dataset, a survey conducted by US Energy Information Administration (EIA). As the MDCEV model is a non-linear framework, independent variable impacts are not readily available from the model results. Hence, to further augment the value of the proposed framework, we conduct a comprehensive policy analysis exercise to illustrate the sensitivity of energy usage choice to various independent variables from the model.

3. Modelling approach

The total energy use of a building i depends on choice of the energy type j and the use of each energy type t_j for energy type j and $\sum_{j=1}^{J} t_{ij} = T_i$. From here on, we suppress the index i in the expressions for ease of presentation. The utility (U_i) derived by allocating the total energy use of a residential building among j energy sources, $T = \{t_1, t_2, t_3, \cdots\cdots, t_J\}$ is given as

$$U_i = \sum_{j=1}^{J} \gamma_j \psi_j exp(\varepsilon_j) ln\left(\frac{t_j}{\gamma_j} + 1\right) \quad (1)$$

ψ_j represents the baseline marginal utility of energy source j's utilization. The usage of energy source j depends on the consumption of j by various end-use utility types. So ψ_j is parameterized as $\psi_j = \exp(\beta_j x_j); \beta_j$ represents a vector of parameters and x_j represents a vector of exogeneous variables influencing the usage of energy alternative j. γ_j is the translation parameter, which also serves to define the satiation effect. ε_j is the stochastic error component that captures the unobserved component of baseline utility.

The above optimization problem can be solved by forming Lagrangian function for the usage constraint and subsequently applying Kuhn-Tucker first order conditions (similar to [48]). In the above model structure, we assume ε_j to be standard extreme value distribution and are independently and identically distributed across the alternatives to derive a closed form of probability expression [49].

$$P(T) = \left[\prod_{l=1}^{m} C_l\right]\left[\sum_{l=1}^{m} \frac{1}{C_l}\right]\left[\frac{\prod_{l=1}^{m} e^{-v_l}}{\left[\sum_{j=1}^{J} e^{-v_j}\right]^{m_j}}\right](m-1)! \cdot \forall j, if\,(t_j > 0) \quad (2)$$

Where $C_j = \frac{1}{t_j + \gamma_j}$

In the above probability expression, m is the number of energy alternatives with non-zero use ($t_j > 0$) and $v_j = \beta_j x_j$.

4. Data

Residential Energy Consumption Survey (RECS) is conducted by EIA USA. EIA conducts survey every 5 years, first of its kind was conducted in 1978. The current research uses the recent RECS survey dataset conducted in 2015. The survey data was collected from 5600 households selected at random using a complex multistage, area-probability sample design, which represents 118.2 million US household (see EIA 2019 [50]). From the data, 4000 records were randomly sampled for estimation and remaining records were set aside for validation.

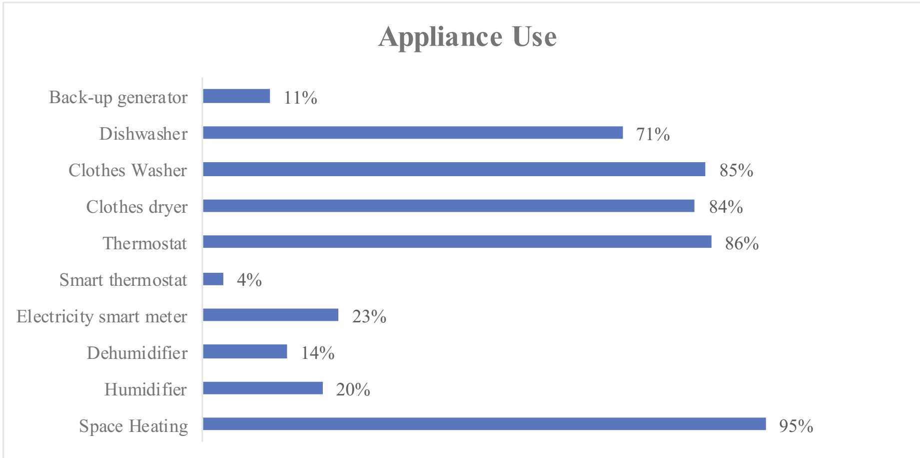

Fig. 2. Share of various appliances used in HH.

In the RECS dataset, total energy use of residential buildings is presented in British Thermal Units[1] (BTUs). The total energy use of each building is provided for electricity, natural gas, fuel oil and propane (LPG). Electricity is available to all the households and natural gas is available only to a section of the respondents. Availability information for fuel oil and LPG is not available in the dataset. Electricity is consumed by all the households, whereas the other energy sources are not consumed by all the buildings. Table 2 presents the descriptive statistics of usage of energy alternatives. Average usage of these energy alternatives by the households using them is presented in the dataset.

From Table 2, it can be observed that all the buildings consume electricity whereas the other energy sources are not consumed by all the other sources. So, the MDC scenario of selection of energy source and its use has electricity as an outside good (for detailed explanation of outside good see [49,51]).

Along with the usage of energy alternatives, RECS dataset provides the details about location characteristics, household characteristics, housing characteristics, appliance use, and climatic variables. Fig. 1 presents descriptive statistics of characteristics of survey respondent and location variables. From the table we can see that 14% of the responses are from the dwellings in the north east census region, around 23% are from the dwellings in the mid-west region, 27% are from the dwellings in south region and 6% are from the dwellings in the west. Out of the sample, a majority of the households are "White Alone" households with a share more than 82% and African Americans represent around 10% of the survey responses. Among these responses 4% of the respondents are the Asians. In the responses, a majority of the respondents are associate degree holders (33%) and the respondents with at least an undergraduate degree is 36%. Household income is also presented in RECS as a categorical variable. The distribution of household income reveals that around 40% of the responses have income less than 40,000$, 29% of the individuals have income between 40,000 to 80,000$ and the rest (31%) has income more than 80,000$.

Various appliances are used in each household. Fig. 2 represents the share of respondents using those various appliances. In the survey, more than 75% of the responses use dish washer, washer, drier, space heating and thermostats for heating. Less than 25% of the households have electric smart meter, humidifier and dehumidifiers. Fig. 3 provides the shares of various types of housing characteristics such as year of construction of the building, type of housing, number of stories of the building, type of occupancy and other amenities. Table 3 presents descriptives of various other continuous variables present in the dataset including number of appliances (such as different types of fans, televisions, smart phones, light bulbs), number of rooms in the dwelling (bedrooms, bathrooms, total rooms), number of household members (adult, children and total), and square footage of building (for heating, cooling and total floor area). Finally, for climatic variables, Heating Degree Days (HDD) and Cooling Degree Days (CDD) are considered to quantify the demand for energy needed for heating and cooling requirements of a building respectively. It is defined as the number of degrees that a day's average temperature is below (above) 65°F for HDD (CDD).

5. Model results

The MDCEV model estimation was based on removing the statistically insignificant variables in a systematic process based on statistical confidence (95% confidence level) and parameter interpretability. The specification process was also guided by prior research and parsimony considerations. The estimates of best model fit are presented in Table 4. The MDCEV model results can be interpreted as follows: a positive (negative) sign of the estimate indicates that the probability of usage of the alternative increases (decreases) with the increase in the variable. In the ensuing discussion, the model estimates from the model are discussed by variable groups: (a) location characteristics, (b) household characteristics (c) housing characteristics and (d) appliance use variables and (e) climatic variables.

Constants: The baseline constants do not provide any interpretations, after introduction of other independent variables.

Location characteristics: Among the location characteristics explored, census region of the residential unit and location context classified as rural or urban offered significant results. Census region variable in the energy use model provides the inherent regional differences in energy selection and usage across the country. From the model results, relative to West census region, we observe that LPG and fuel oil are preferred in the north east region while natural gas is less preferred. In the mid-west region, all non-electricity energy sources are less preferred relative to the West region. In the south census region, a preference for fuel oil is observed. In terms of location context, we find that units in rural locations have a preferences for LPG and fuel oil (see [36]

[1] The British Thermal Unit (BTU) is defined as the amount of heat required to raise the temperature of one pound of water by 1° Fahrenheit, which is equivalent to 1055 Joules (in SI units).

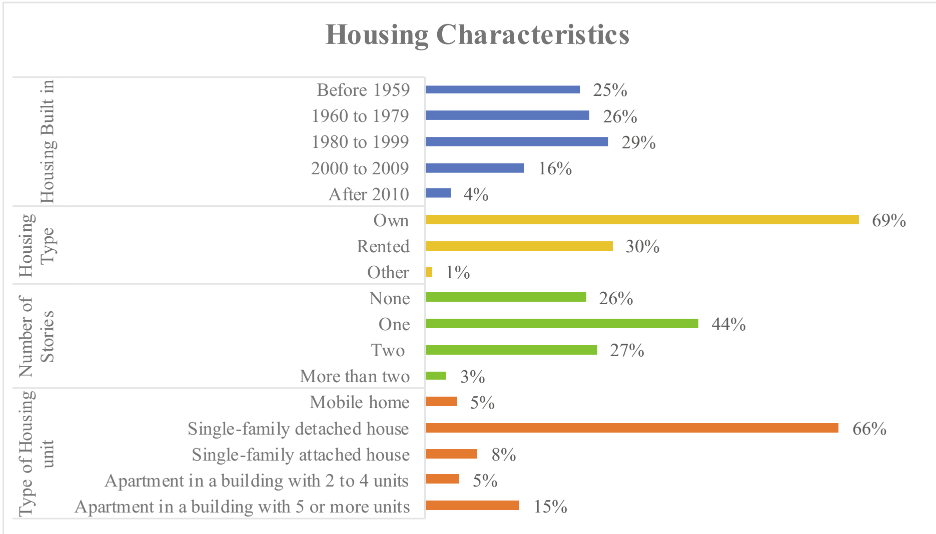

Fig. 3. Descriptives of various housing characteristics.

Table 3
Descriptive statistics of other continuous variables.

Variables	Minimum	Maximum	Mean
Number of bedrooms	0	10	2.83
Number of full bathrooms	0	6	1.75
Number of half bathrooms	0	4	0.32
Number of other rooms	1	14	3.36
Total number of rooms in the housing unit, excluding bathrooms	1	19	6.19
Number of televisions used	0	9	2.36
Number of smart phones	0	8	1.61
Number of ceiling fans used	0	14	2.24
Number of floor, window, or table fans used	0	14	0.82
Number of whole house fans used	0	9	0.08
Number of light bulbs installed inside the home	1	5	2.07
Number of inside light bulbs turned on at least 4 h a day	0	80	7.34
Number of household members	1	12	2.58
Number of household members age 18 or older	1	10	1.97
Number of household members age 17 or younger	0	10	0.61
Number of weekdays someone is at home	0	5	3.4
Total cooled square footage	0	8066	1454.52
Total heated square footage	0	8066	1815.81
Total square footage (used for publication)	221	8501	2081.44
CDD in 2015, base temperature 65F[1]	0	6607	1719.21
HDD in 2015, base temperature 65F	0	9843	3707.85

[1] The unit of Heating Degree Days (HDD) and Cooling Degree Days (CDD) is degree days (DD).

for similar results). From these results, it is apparent that residential energy use in rural areas is reliant on LPG and Fuel oil. To reduce their greenhouse gas (GHG) emission burden it might be prudent to encourage policies incentivizing solar energy adoption in these regions with a particular focus in the South and West regions.

Household characteristics: Household income has a perceptible influence on energy mix. Specifically, we find that households with annual income under 80 k or between 80 and 100 k have lower inclination for natural gas and LPG. On the other hand, households with income greater than 100 k exhibit preference for fuel oil. In addition to income, household size and number of children variables offer expected results. In particular, increased number of individuals lead to higher electricity, natural gas and fuel oil consumption. Also, it is interesting to note that children contribute more to energy use than adults in the household (as indicated by the positive sign for the number of children parameter). These results highlight the key differences in residential energy use across households with varying income and household composition.

Housing characteristics: In terms of Housing type, relative to all other categories, mobile home and apartment categories offer significant results. Specifically, mobile home units have a higher preference for LPG and fuel oil while apartments are likely to exhibit higher preference for fuel oil. The building construction period presents interesting relationships with energy choice. Relative to houses constructed before 1960, houses from 1960 have higher inclination for electricity, natural gas and fuel oil. The construction time period between 1980 and 2000 indicates a higher preference for LPG energy choice. This is the only time period of housing construction where electricity is less preferred for usage. In subsequent time periods, while LPG is preferred to natural gas and fuel oil, electricity is again the most preferred energy source. As the square footage of the house increases (considered in the form of a natural logarithm), we observe that preference for non-electricity energy sources increases indicating that larger houses are more likely to have energy mix from multiple sources (see [36]). In addition to the overall square footage, we also explored the impact of number of rooms and number of bedrooms on energy use. As expected, these variables are associated with increased electricity use (relative to LPG use).

The number of rooms is associated with a reduced use of natural gas and increased use of fuel oil. The number of bedrooms also is associated with increased use of natural gas and fuel oil. It is important to note here that square footage, number of rooms and number of bedrooms are variables that influence each other (yet not correlated). So, their impact on energy selection and usage needs to be considered together.

Appliance use variables: The RECS dataset provides the information of various appliances used in the household. In the MDCEV model, we tested for the impact of these appliances on energy usage. The findings are intuitive. The presence and use of the appliances result in increased electricity usage (such as for refrigerator,

Table 4
Results of MDCEV model estimation.

Variables	Electricity	Natural Gas	LPG	Fuel Oil
	Coefficient (t-statistic)	Coefficient (t-statistic)	Coefficient (t-statistic)	Coefficient (t-statistic)
Constant	–	−5.777 (−866.50)	−6.348 (−5 0 9)	−36.465 (−246.72)
Location Characteristics				
Region (West census region is the base)				
North-east	–	−0.470 (−654.70)	0.545 (352.35)	2.626 (706.35)
Mid-west	–	−0.118 (−176.94)	−0.192 (−122.38)	−1.129 (−244.74)
South	–	−1.159 (−1874.95)	−0.210 (−137.29)	1.651 (434.59)
Rural region	–	−1.532 (−2246.49)	1.865 (1968.85)	0.486 (374.96)
Household Characteristics				
Income				
<80 k	–	−0.273 (−456.78)	−0.263 (−226.36)	–
80 to 100 k	–	−0.088 (−128.33)	−0.263 (−226.36)	–
>100 k	–	–	–	0.024 (15.67)
HH size	0.152 (280.11)	0.065 (116.92)	–	0.174 (227.21)
Number of children	0.064 (89.00)	0.113 (152.75)	–	0.064 (62.45)
Housing Characteristics				
Housing type (Other housing types are base)				
Mobile home	–	−0.693 (−564.01)	0.463 (312.13)	0.187 (65.90)
Apartment	–	−0.337 (−400.82)	−0.38 (−145.28)	0.943 (402.00)
Building construction period (Before 1960 is base)				
1960 to 1980	0.388 (300.62)	0.139 (103.12)	–	0.105 (57.67)
1980 to 2000	−0.176 (−156.23)	−0.431 (−361.80)	–	−1.366 (−578.04)
2000 to 2010	0.057 (45.96)	−0.285 (−219.21)	–	−1.100 (−451.46)
After 2010	0.203 (89.08)	−0.091 (−37.13)	–	−1.663 (−237.4)
Log (Square footage)	–	0.210 (162.73)	0.845 (334.33)	0.711 (199.98)
Number of bedrooms	0.088 (132.55)	0.168 (243.21)	–	0.032 (32.61)
Total number of rooms	0.024 (88.70)	−0.013 (−43.61)	–	0.01 (21.92)
Appliance Use				
Backup generator	−0.427 (−308.18)	−0.800 (−549.96)	0.228 (145.53)	–
Electricity generated from solar	0.294 (194.47)	–	–	–
Number of refrigerators	0.079 (257.90)	–	–	–
Dryer used	0.358 (558.92)	–	–	–
Outside grill used	0.020 (43.47)	–	–	–
Number of color TVs	0.043 (252.56)	–	–	–
Number of play stations	0.057 (254.25)	–	–	–
Coffee maker	0.065 (162.73)	–	–	–
Crockpot used	0.004 (9.58)	–	–	–
Other appliances	0.118 (194.83)	–	–	–
Number of desktops	0.006 (22.65)	–	–	–
Number of smart phones	0.009 (46.56)	–	–	–
Internet used	0.125 (210.64)	–	–	–
Smart meter for electricity	−0.093 (−213.31)	–	–	–
Space heating used	−4.302 (−30.58)	−4.520 (−32.13)	−5.162 (−36.69)	–
AC used	0.534 (420.97)	0.245 (186.97)	0.015 (11.44)	–
Humidifier used	0.257 (205.49)	0.477 (374.29)	0.218 (142.03)	–
Climatic Variables				
Log (HDD)	–	0.938 (947.90)	0.340 (159.09)	6.002 (652.62)
Log (CDD)	–	0.274 (259.34)	−0.360 (−217.25)	0.458 (101.32)
Satiation (Gamma) parameters	–	71.999 (1752.78)	28.087 (1200.63)	366.048 (343.51)
Total weighted Log-likelihood at convergence	−324,489,600			
Total weighted Log-likelihood at constants only	−349,548,000			

dryer, grill, televisions, play stations, coffee maker, crockpot, desktops, smartphones). Interestingly we find that units with backup generator have lower electricity and natural gas use while LPG usage is likely to be higher. In units with solar electricity production, we observe a higher electricity usage. This could potentially indicate the reason for solar installation (to reduce electricity costs). Another interesting finding pertains to internet usage. Households with internet are likely to use more electricity. In households with smart meters there is a reduced use of electricity. The adoption of space heating is associated with lower electricity, natural gas and LPG usage indicating a preference for fuel oil. AC and humidifier usage are associated with higher usage for all energy sources (except fuel oil).

Based on these results, several important recommendations can be made. It might be useful for utility providers to replace existing meters with smart meters to better manage electricity use (see [52]). Also, educating and promoting the increased adoption of space heating infrastructure during excessively cold time periods might present a mechanism to efficiently manage energy use (for example see adoption of solar chimneys [53]). Finally, in residential units with high AC and heating energy usage, providing incentives for retrofitting with smart thermostats and/or insulation retrofit might contribute to greater energy use savings (see [54,55]).

Climatic variables: Log transformed variables of HDD and CDD are used as independent variables in our MDCEV model. The findings indicate that HDD is associated with higher usage for non-electric energy sources potentially alluding to these energy sources employed for heating. In terms of CDD, we find that while natural gas usage is positively affected, LPG and fuel oil usage is negatively affected. The residential buildings in the regions where the climatic conditions are different from optimal climate from energy use perspective (mild weather), implementing periodic insulation checks, offering incentives for retrofitting and optimizing HVAC design can potentially reduce energy needs for heating and cooling requirements [56].

Satiation (gamma) parameter: The (γ_j)parameter is only estimated for alternatives that have non-zero usage possibility (for more information about the corner solutions see [49,51]). From our results, we find that fuel oil has the highest satiation indicating that marginal utility for energy usage drops rapidly thus resulting in smaller usage levels. Between natural gas and LPG, natural gas has a higher satiation.

Validation: The RECS data that is not used in model estimation (1686 records) is used to validate the model fit of the estimated model. The log-likelihood is estimated on the validation sample using the model estimates discussed earlier. While there is no way to compare the total log-likelihood functions, we can compare the average record level log-likelihood and adjusted ρ^2 to examine if there are major differences in the predictions between estimation and validation sample. The average log-likelihood (adjusted ρ^2) of the validation sample using the estimated model is -83318.6 (0.064) while the corresponding value of the estimation sample is -81122.4 (0.072). The difference in the two measures is reasonable and does not appear to reflect any over-fitting of the model in the estimation sample.

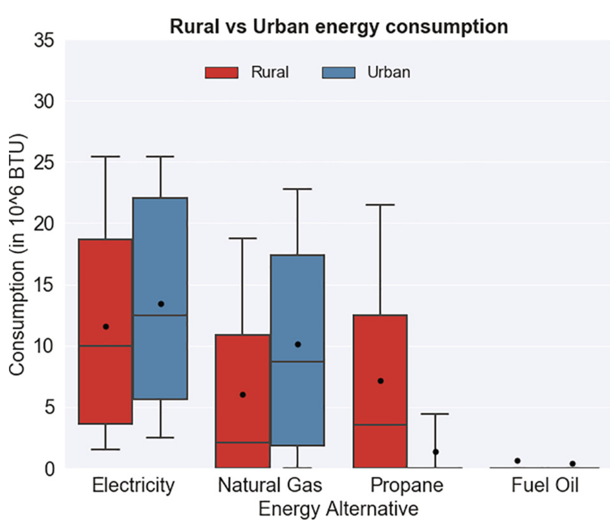

Fig. 4. Energy use in rural vs urban locations.

6. Effect of various attributes on energy use

The model estimates for the MDCEV model directly do not provide the marginal impacts of the independent variables. Towards this end, we predict the energy use behavior in residential units due to changes in the independent variables significant in the MDCEV model. The prediction follows forecasting method for Kuhn-Tucker consumer demand model system developed by Pinjari and Bhat (see [36] for more details). The reader would note the energy predictions are generated using multiple realizations of the error terms to account for stochasticity in the prediction process. For this prediction process, we have selected a record in the dataset that is in Southern census region, owning a housing unit built between 1980 and 2000, with a HDD of 3900, CDD of 1330, with household income ranging between 40,000 to 60,000$ and uses basic appliances (namely; backup electricity generation, refrigerator, space heating, cooling, stove, TV, washing machine and drier), listed in the MDCEV model. The energy use is forecasted by using all the parameters to compare the energy use by location (rural or urban), HH size, number of children, size of the residential unit (number of bedrooms and area of the unit) are studied. The process can be extended to any household unit of interest.

Rural vs *Urban:* In this, we specifically compared residential energy use of various sources of energy in a rural location and urban location, while all the other variables remained constant. The variation of electricity, natural gas, propane and fuel oil use are presented in Fig. 4. From the figure, usage of electricity, natural gas is likely to be higher in the urban region, while propane is more likely to be used in rural region. The fuel oil use is almost negligible for the chosen household unit.

Household size: In this, we specifically studied the change in residential energy use with increase in household size. Given the differences observed for adults and children, we study energy predictions separately for them. To understand the impact of adults, we fixed the number of children to 0. For studying the impact of children, we fixed the number of adults to two. The prediction results are presented in Fig. 5. The figures clearly illustrate how energy usage by source increases with additional household members.

Size of housing unit: In the MDCEV model formulation, various housing unit characteristics were explored. They are size of the

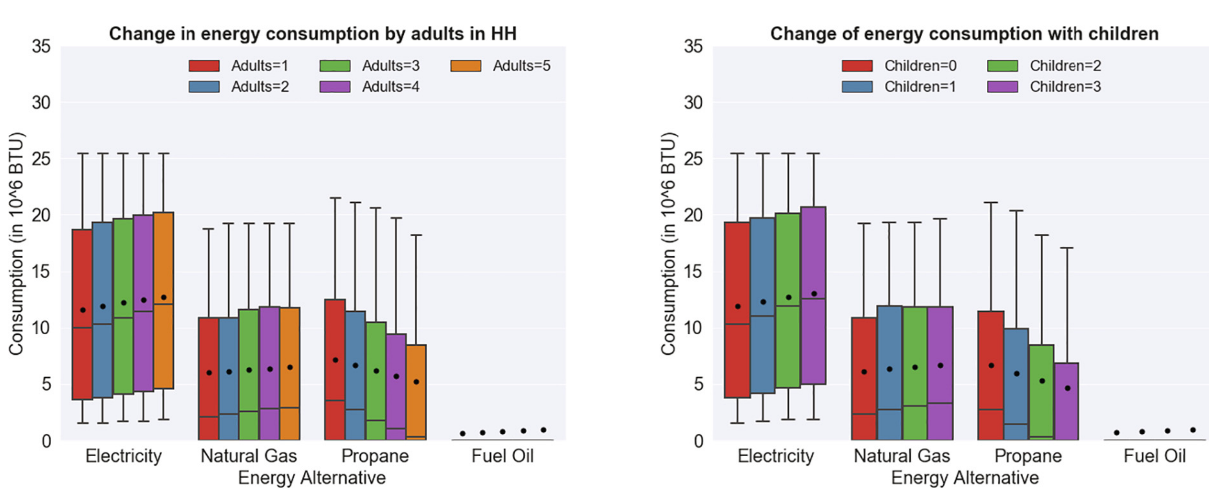

Fig. 5. Variation of energy use with HH size.

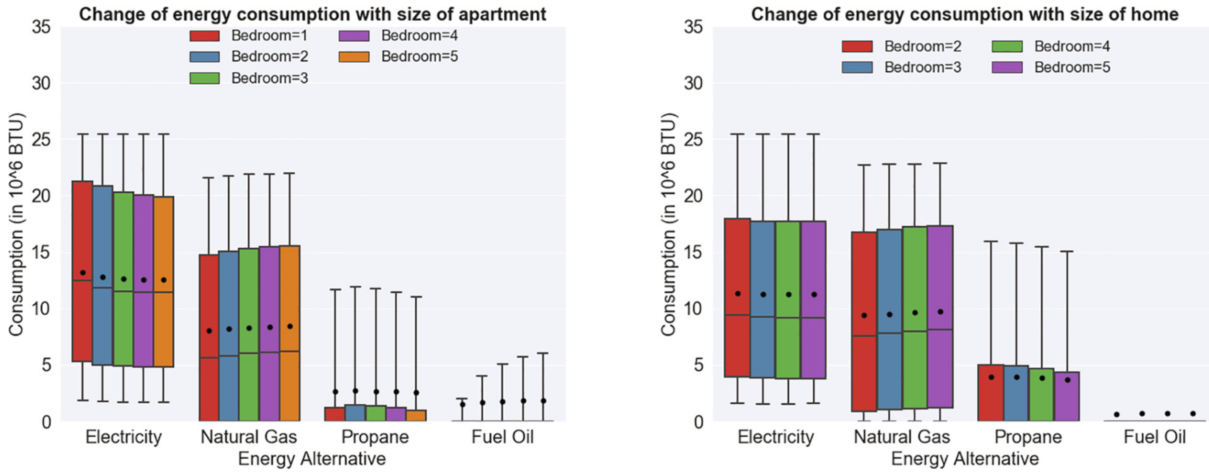

Fig. 6. Variation of energy use with housing unit size.

unit (in log (square footage)), number of bedrooms and total number of rooms in the unit. These characteristics are interdependent on each other. Along with these characteristics, the type of housing also has a great influence on the energy use of the unit. In the MDCEV model, the estimate for the type of housing unit (apartment or mobile home) is estimated by considering single family housing as the base. So, the housing unit characteristics are studied separately for apartment and single-family housing. In the effect of housing size in the context of apartment type housing, the base case is a single bedroom apartment of size 500 sqft, with a total of 3 rooms in it. The energy use is studied with increase of one bedroom and one bathroom at a time, of size 300 sqft. Similarly, in the context of single-family housing, the base case is a double bedroom house of size 1200 sqft, with a total of 4 rooms in it. The energy use is studied with increase of one bedroom and one bathroom at a time, of size 400 sqft. Fig. 6 presents the variation of usage of energy sources with change in size of the apartment or individual home. From the figure, we can observe that the electricity use initially tends to decrease with increase in size of apartment (or single-family house) but increases gradually. In context of natural gas, the energy use constantly increases with increase in size of the apartment (or single-family house). While there is no noticeable change in usage of propane or fuel oil.

7. Summary and conclusions

In this paper, we focus on developing a residential energy use prediction framework for United States using the 2015 Residential Energy Consumption Survey (RECS) that provides energy use details for electricity, natural gas, fuel oil and Liquefied Petroleum Gas (LPG). Residential energy usage includes addressing two decisions: (1) source of energy (such as electric and natural gas) and (2) usage by energy source for various purposes. Towards studying the energy usage process, a MDCEV model is estimated by using an exhaustive set of independent variables including location characteristics, household characteristics, housing characteristics, appliance use and climatic characteristics.

The MDCEV model revealed the inherent regional differences (by census region) in energy selection and usage across the country. It is interesting to find that units in rural locations prefer LPG and fuel oil. Household characteristics, such as income and household configuration (size and number of children), influence the energy mix. It is interesting to find that children contribute more to energy use than adults in the household. Various housing characteristics such as type of housing unit, construction period, size and configuration of the housing unit have an impact on energy mix and usage. The influence of various appliances used in the household are studied. The presence and use of the appliances result in increased electricity usage (such as for refrigerator, dryer, grill, televisions, play stations, coffee maker, crockpot, desktops, smartphones, AC, humidifier). Interestingly we found that units with backup generator have lower electricity and natural gas use while LPG usage is likely to be higher. The adoption of space heating is associated with lower electricity, natural gas and LPG usage indicating a preference for fuel oil. The findings indicate that HDD is associated with higher usage for non-electric energy sources. In terms of CDD, we find that while natural gas usage is positively affected, LPG and fuel oil usage is negatively affected. Several important policy recommendations can be made based on the model results. In rural areas, to shift the energy mix toward renewable energy, incentives for solar installations might be appropriate. For local utility providers, important recommendations include updating the metering systems to smart meters, offering periodic insulation checks and reduced cost retrofitting, retrofitting with smart thermostats and educating consumers about the benefits of space heating can provide reductions in overall energy use.

To illustrate the applicability of the proposed model in prediction, we conducted sensitivity analysis of energy mix and usage using a specific residential unit from the dataset. The energy use is forecasted to compare the usage by location (rural or urban), HH size, number of children, size of the residential unit (number of bedrooms and area of the unit). The prediction analysis provides a representation of how various variables considered in the policy analysis affect energy mix and usage. The approach presented can be extended to draw insights for any household unit.

The results from the proposed research will also be useful for establishing a base line of residential energy use. The model developed can be employed to build an energy demand simulation for urban regions by employing synthetic population generation [57]. Using synthetic population generation, urban household population can be synthesized and their energy demand can be predicted using our model system. The prediction will serve as an annual demand profile for residential energy use and present utility providers with an expected grid demand. With the advent of electric vehicles, a newer component of energy use will emerge and is likely to alter the current energy use patterns. The model developed will allow us to provide the non-electric vehicle energy use for urban regions. The baseline residential energy use patterns can be appropriately augmented with expected household level electric vehicle energy demand to obtain overall energy demand on the grid under evolving future vehicle mix scenarios. The proposed model system can also be applied to study energy use across commercial energy sector using appropriate data (for example see sample results from the analysis of Commercial Building Energy Consumption Survey (CBECS) data). Finally, building on the current study future research efforts can also consider energy use analysis at a fine resolution such as energy use by activity (lighting, heating and so on). The MDCEV model from our research can be easily extended to conduct data analysis of energy use by energy use activities.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors acknowledge the data from RECS survey and would like to thank EIA for providing the survey data on their website. We also thank Dr. Chandra Bhat for providing MDCEV model code on his research website.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.enbuild.2021.110934.

References

[1] US-EIA. What is the United States' share of world energy consumption? https://www.eia.gov/tools/faqs/faq.php?id=87&t=1. Accessed 15th March, 2019.
[2] Worldometer. U.S. Population (Live). http://www.worldometers.info/world-population/us-population/. Accessed March 15, 2019.
[3] US-DOE. Residential Program Solution Center - Energy Data Facts. https://rpsc.energy.gov/energy-data-facts. Accessed 15 September, 2019.
[4] IEA. International Energy Agency - World Energy Balances: Overview, 2017. https://www.iea.org/publications/freepublications/publication/WorldEnergyBalances2017Overview.pdf. Accessed 5th Septmenber, 2019.
[5] P. Nejat, F. Jomehzadeh, M.M. Taheri, M. Gohari, M.Z.A. Majid, A global review of energy consumption, CO2 emissions and policy in the residential sector (with an overview of the top ten CO2 emitting countries), Renew. Sustain. Energy Rev. 43 (2015) 843–862.
[6] A. Sobhani, N. Eluru, A. Faghih-Imani, A latent segmentation based multiple discrete continuous extreme value model, Transp. Res. Part B: Methodol. 58 (2013) 154–169.
[7] C.R. Bhat, M. Castro, M. Khan, A new estimation approach for the multiple discrete–continuous probit (MDCP) choice model, Transp. Res. Part B: Methodol. 55 (2013) 1–22.
[8] M. Castro, C.R. Bhat, R.M. Pendyala, S.R. Jara-Díaz, Accommodating multiple constraints in the multiple discrete–continuous extreme value (MDCEV) choice model, Transp. Res. Part B: Methodol. 46 (6) (2012) 729–743.
[9] K. Amasyali, N.M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renew. Sustain. Energy Rev. 81 (2018) 1192–1205.
[10] Y. Wei, X. Zhang, Y. Shi, L. Xia, S. Pan, J. Wu, M. Han, X. Zhao, A review of data-driven approaches for prediction and classification of building energy consumption, Renew. Sustain. Energy Rev. 82 (2018) 1027–1047.
[11] M. Azaza, F. Wallin, Smart meter data clustering using consumption indicators: responsibility factor and consumption variability, Energy Procedia 142 (2017) 2236–2242.
[12] A. Al-Wakeel, J. Wu, N. Jenkins, K-means based load estimation of domestic smart meter measurements, Appl. Energy 194 (2017) 333–342.
[13] I. Benítez, J.-L. Díez, A. Quijano, I. Delgado, Dynamic clustering of residential electricity consumption time series data based on Hausdorff distance, Electr. Power Syst. Res. 140 (2016) 517–526.
[14] F. Belaid, T. Garcia, Understanding the spectrum of residential energy-saving behaviours: French evidence using disaggregated data, Energy Econ. 57 (2016) 204–214.
[15] O.I. Asensio, M.A. Delmas, The dynamics of behavior change: evidence from energy conservation, J. Econ. Behav. Organ. 126 (2016) 196–212.
[16] B. Anderson, S. Lin, A. Newing, A. Bahaj, P. James, Electricity consumption and household characteristics: implications for census-taking in a smart metered future, Comput. Environ. Urban Syst. 63 (2017) 58–67.
[17] M. Bedir, E. Hasselaar, L. Itard, Determinants of electricity consumption in Dutch dwellings, Energy Build. 58 (2013) 194–207.
[18] L. Blázquez, N. Boogen, M. Filippini, Residential electricity demand in Spain: new empirical evidence using aggregate data, Energy Econ. 36 (2013) 648–657.
[19] H.S. Boudet, J.A. Flora, K.C. Armel, Clustering household energy-saving behaviours by behavioural attribute, Energy Policy 92 (2016) 444–454.
[20] D. Brounen, N. Kok, J.M. Quigley, Residential energy use and conservation: Economics and demographics, Eur. Econ. Rev. 56 (5) (2012) 931–945.
[21] J. Szoplik, Forecasting of natural gas consumption with artificial neural networks, Energy 85 (2015) 208–220.
[22] P. Potočnik, B. Soldo, G. Šimunović, T. Šarić, A. Jeromen, E. Govekar, Comparison of static and adaptive models for short-term residential natural gas forecasting in Croatia, Appl. Energy 129 (2014) 94–103.
[23] F. Taşpınar, N. Celebi, N. Tutkun, Forecasting of daily natural gas consumption on regional basis in Turkey using various computational methods, Energy Build. 56 (2013) 23–31.
[24] M. Bilgili, B. Sahin, A. Yasar, E. Simsek, Electric energy demands of Turkey in residential and industrial sectors, Renew. Sustain. Energy Rev. 16 (1) (2012) 404–414.
[25] R.E. Edwards, J. New, L.E. Parker, Predicting future hourly residential electrical consumption: a machine learning case study, Energy Build. 49 (2012) 591–603.
[26] L.C. Hunt, G. Judge, Y. Ninomiya, Underlying trends and seasonality in UK energy demand: a sectoral analysis, Energy Econ. 25 (1) (2003) 93–118.
[27] T.N. Atalla, L.C. Hunt, Modelling residential electricity demand in the GCC countries, Energy Econ. 59 (2016) 149–158.
[28] Z. Dilaver, L.C. Hunt, Modelling and forecasting Turkish residential electricity demand, Energy Policy 39 (6) (2011) 3117–3127.
[29] M. Forouzanfar, A. Doustmohammadi, M.B. Menhaj, S. Hasanzadeh, Modeling and estimation of the natural gas consumption for residential and commercial sectors in Iran, Appl. Energy 87 (1) (2010) 268–274.
[30] B.C. Ervural, O.F. Beyca, S. Zaim, Model estimation of ARMA using genetic algorithms: A case study of forecasting natural gas consumption, Procedia-Soc. Behav. Sci. 235 (2016) 537–545.
[31] R.V. Jones, K.J. Lomas, Determinants of high electrical energy demand in UK homes: Socio-economic and dwelling characteristics, Energy Build. 101 (2015) 24–34.
[32] A. Kavousian, R. Rajagopal, M. Fischer, Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior, Energy 55 (2013) 184–194.
[33] G. Huebner, D. Shipworth, I. Hamilton, Z. Chalabi, T. Oreszczyn, Understanding electricity consumption: A comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes, Appl. Energy 177 (2016) 692–702.
[34] S. Hori, K. Kondo, D. Nogata, H. Ben, The determinants of household energy-saving behavior: survey and comparison in five major Asian cities, Energy Policy 52 (2013) 354–362.
[35] W.-H. Huang, The determinants of household electricity consumption in Taiwan: evidence from quantile regression, Energy 87 (2015) 120–133.
[36] A.R. Pinjari, C.R. Bhat. Computationally efficient forecasting procedures for Kuhn-Tucker consumer demand model systems: application to residential energy consumption analysis. In, 2011.
[37] D. Wiesmann, I.L. Azevedo, P. Ferrão, J.E. Fernández, Residential electricity consumption in Portugal: Findings from top-down and bottom-up models, Energy Policy 39 (5) (2011) 2772–2779.
[38] B. Yu, J. Zhang, Modeling household energy consumption behavior: a comparative analysis, Transp. Res. Part D: Transp. Environ. 39 (2015) 126–140.
[39] L. Dale, S. Fujita, F. Vasquez, M. Moezzi, M. Hanemann, S. Guerrero, L. Lutzenhiser. Price impact on the demand for water and energy in California Residences. California Climate Change Center, 2009.
[40] J.A. Dubin, D.L. McFadden. An econometric analysis of residential electric appliance holdings and consumption. Econometrica: J. Econ. Soc. 1984, pp. 345-362.
[41] M. Filippini, S. Pachauri, Elasticities of electricity demand in urban Indian households, Energy Policy 32 (3) (2004) 429–436.
[42] E.T. Mansur, R. Mendelsohn, W. Morrison, Climate change adaptation: a study of fuel choice and consumption in the US energy sector, J. Environ. Econ. Manage. 55 (2) (2008) 175–193.
[43] P.K. Narayan, R. Smyth, The residential demand for electricity in Australia: an application of the bounds testing approach to cointegration, Energy Policy 33 (4) (2005) 467–474.
[44] R. Nesbakken, Energy consumption for space heating: a discrete–continuous approach, Scand. J. Econ. 103 (1) (2001) 165–184.
[45] K. Vaage, Heating technology and energy use: a discrete/continuous choice approach to Norwegian household energy demand, Energy Econ. 22 (6) (2000) 649–666.
[46] D.J. Sailor, J.R. Muñoz, Sensitivity of electricity and natural gas consumption to climate in the USA—methodology and results for eight states, Energy 22 (10) (1997) 987–998.

[47] J. Harold, S. Lyons, J. Cullinan, The determinants of residential gas demand in Ireland, Energy Econ. 51 (2015) 475–483.
[48] T.J. Wales, A.D. Woodland, Estimation of consumer demand systems with binding non-negativity constraints, J. Econometrics 21 (3) (1983) 263–285.
[49] C.R. Bhat, A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions, Transp. Res. Part B: Methodol. 39 (8) (2005) 679–707.
[50] EIA. Residential Energy Consumption Survey (RECS). https://www.eia.gov/consumption/residential/. Accessed 15th September, 2019.
[51] C.R. Bhat, The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions, Transp. Res. Part B: Methodol. 42 (3) (2008) 274–303.
[52] Q. Sun, H. Li, Z. Ma, C. Wang, J. Campillo, Q. Zhang, F. Wallin, J. Guo, A comprehensive review of smart energy meters in intelligent energy networks, IEEE Internet Things J. 3 (4) (2015) 464–479.
[53] N. Monghasemi, A. Vadiee, A review of solar chimney integrated systems for space heating and cooling application, Renew. Sustain. Energy Rev. 81 (2018) 2714–2730.
[54] S. Oh, J.S. Haberl, J.-C. Baltazar, Analysis of zone-by-zone indoor environmental conditions and electricity savings from the use of a smart thermostat: a residential case study, Sci. Technol. Built Environ. 26 (3) (2020) 285–303.
[55] S. Chidiac, E. Catania, E. Morofsky, S. Foo, A screening methodology for implementing cost effective energy retrofit measures in Canadian office buildings, Energy Build. 43 (2–3) (2011) 614–620.
[56] A. Higgins, M. Syme, J. McGregor, L. Marquez, S. Seo, Forecasting uptake of retrofit packages in office building stock under government incentives, Energy Policy 65 (2014) 501–511.
[57] K.C. Konduri, D. You, V.M. Garikapati, R.M. Pendyala, Enhanced synthetic population generator that accommodates control variables at multiple geographic resolutions, Transp. Res. Rec. 2563 (1) (2016) 40–50.