Energy 66 (2014) 722–731

Contents lists available at ScienceDirect

Energy

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

A multilevel regression approach to understand effects of environment indicators and household features on residential energy consumption

Geoffrey K.F. Tso*, Jingjing Guan

Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 26 June 2013
Received in revised form
13 January 2014
Accepted 16 January 2014
Available online 22 February 2014

Keywords:
Residential energy consumption
Multilevel regression
Area variations

ABSTRACT

Modeling residential energy consumption survey (RECS) data is a complex socio-technical problem that involves macroeconomics, climate, physical characteristics of housing, household demographics and usage of appliances. A multilevel regression (MR) model is introduced to calculate the magnitude and significance of effects of environment indicators and household features on residential energy consumption (REC). MR helps construct a conceptual framework and organize explanatory variables. The benefit of this approach is that based on stratified sampling schemes, MR extracts area effects from total variations of REC and explains the remaining variations with manifest variables and their interactions. Using the US 2009 RECS micro data consisting of 10,838 unique cases, 26 primary determinants of REC are found to be division groups, housing type, house size, usage of space heating equipment, household size and use of air-conditioning, etc. MR helps to quantify 82% of area effects and 47% of household effects. Proportion of the overall explained variance proportion is 53% compared to <40% using OLS regression models.

1. Introduction

The residential sector accounted for 22% of the total energy consumption in the US during 2011. This proportion was less than 10% in the late 1940s. Total residential annual energy consumption increased from 5989 trillion BTU in 1950 to 21,619 trillion BTU in 2011. National and global energy markets are challenged by explosively growing population, fast penetration of modern home appliances, global energy shortages and increasing energy prices. Thus, understanding energy consumption behaviors in the residential sector is important for all other sectors in US. Yet, the Energy Information Administration (EIA) [1] claimed that total U.S. energy consumption in homes has remained relatively stable for years, and the average per-household energy consumption has a decreasing trend. This decrease is due to the increased energy efficiency and offsets the increase in the number and average size of housing units. With a decreasing trend of average household residential energy consumption (REC), it is of interest to explore the latest residential energy consumption survey (RECS) micro dataset in the US.

Besides, due to area variations or regional effects, energy policies such as tier-based policies can easily become ineffective. Regional effects to REC stem from environment, energy and residents. Sailor [2] in 2001 pointed out previous literature exhibited a narrow regional focus and related climate changes to residential electricity consumption of eight states in the US. In the few literature that explicitly consider regional effects on energy consumption, Wang and Wang [3] in 2011 confirmed the existence of regional interaction of biomass consumption in the US by using spatial autoregressive model. Their results show that regional interaction becomes weaker with the farther neighbor states. Thus, we speculate that regional effects might be significant on REC. Besides, regional effects as a key factor of energy plantation, architecture design, city planning and infrastructure, and future energy assessing, do not draw enough attention from REC researchers. It requires sufficient evidence to support governmental policy making, to improve the operations made by industrial practitioners, and to help households understand how they consume energy in homes. Hence, it is of interest to find an advanced statistical modeling approach which can extract area variations and identify key impact factors, to improve modeling approaches of REC.

In this paper, we propose a cross-sectional analysis of household energy consumption in the US by using micro-level data. Instead of

* Corresponding author. Tel: +852 3442 8568; fax: +852 3442 0189.
E-mail address: msgtso@cityu.edu.hk (G.K.F. Tso).

http://dx.doi.org/10.1016/j.energy.2014.01.056
0360-5442/

focusing on residential electricity consumption e.g. [2,4,5] or local area research e.g. [6] or national research without identifying regional effects e.g. [7], we uses a new approach, i.e. a multilevel regression (MR) model, which explicitly considers regional effects within a national sample. The MR model we build is a combination of bottom–up [8,9,10] and top–down model [11,12,13]. It is because we utilize individual household feature data provided by the EIA along with techno-socio-macroeconomic data provided by the BEA. This approach chimes in with the idea that total energy demand can be explained by a wide array of structural indicators, such as the general degree of economic welfare, the extent of electrification, the availability of other energy carriers, the prevalence of energy efficient technologies, the prevailing climate and cultural habits [14–18].

1.1. Contribution

The prime purpose of this paper is to provide a better data-based REC modeling strategy to improve residential energy policy making. We introduce multilevel regression analysis to split total variations of REC among households into area variations and household variations. This is the first known application of MR models for study of REC with the US 2009 RECS micro dataset. Multilevel regression analysis stems from multiple linear regression analysis, which is popular in modeling REC. Multilevel regression analysis can address the problem of clustering of households in modeling REC, which multiple linear regression analysis cannot. Clustering of households has strong impacts on research problems with datasets collected by complex multistage stratified sampling in national surveys. For example, one consequence is that relationships between explanatory variables and response variable are heterogenous among clusters. To handle clustering of households, multilevel regression analysis incorporates information on how consumption disparities are attributed to households and areas they live in. It quantifies the clustering extent of REC among areas, and permits examining cross effects of area-level and household-level factors.

Based on multilevel regression analysis, we propose a conceptual framework for modeling REC. This conceptual framework emphasizes the hierarchical structure within RECS micro dataset. Households are geographically nested into different areas. Households form first hierarchy, whilst areas form second hierarchy. In the context of the US 2009 RECS, as shown in Fig. 1, objects in level 1 are individual households; objects in level 2 of the model refer to clusters of individual households according to their geographical location by divisions or reportable domains, as shown in Figs. 2 and 3. Hence, the data structure brought by multistage stratified sampling schemes is considered. Response variable is REC of individual households. Explanatory variables in the regression model are arranged in accordance with objects that they measure. Level-1 explanatory variables refer to household features, including microclimate, housing construction, socio-demographic and usage of appliances factors. Level-2 explanatory variables are environment indicators, which refer to variables measuring geographical groups of households; they are common to all households living in the same region.

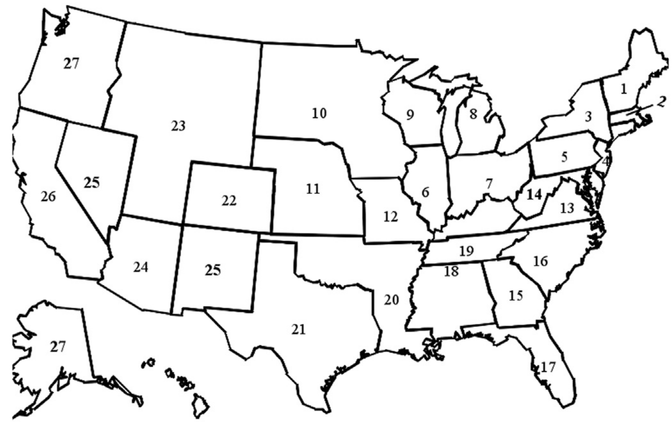

Fig. 2. Reportable domains of US.

We confirm the necessity of using multilevel regression analysis by a finding that clustering effects or regional effects can explain nearly 20% of variances of REC among households. The MR model we have identified can explain 82% of clustering effects, and 47% of variations of household effects. Proportion of the overall explained variance is around 53%. We found 26 significant determinants of REC including division groups, housing type, house size, usage of space heating equipment, household size and use of air-conditioning (AC) etc. Households living in northern parts of the US consume more energy than those living in southern US do. Especially, households living in divisions 1 and 2 consume more energy than others. After ruling out regional effects, ceteris paribus, house size has the largest impact on household energy consumption. For each square feet increase of mean house size, the expected household energy consumption increases by 488,791 kWh/year. Though the highest education of householder, the member(s) who owns the living unit, turns out to be not significant, it mediates house size with a U-shape relationship on REC. We find that single-family detached (SFD) housing is the most energy consuming housing type throughout all divisions in the US. For an extra person living in a house, while compared to divisionwise averages, the expected household energy increases by 219,811 kWh/year. Moreover, program-controlled space heating equipment does help save 11,227 kWh/year compared to those without program control. Basically, our MR model not only avoids biased estimations resulting from traditional linear regression models but also proves that REC disparities among households located in a vast area can be carefully examined by levels of hierarchy.

1.2. Structure of the paper

Section 2 introduces the epistemology of modeling RECS data with multilevel regression analysis. It deliberates on the motivation for applying multilevel regression analysis to explain REC, the advantages of using multilevel regression analysis, and a conceptual framework to analyze the 2009 RECS US micro dataset. The

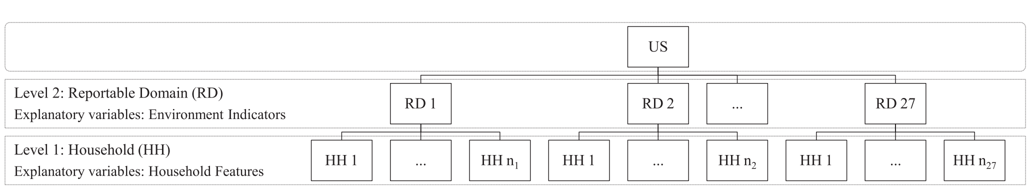

Fig. 1. The conceptual framework of modeling REC with multilevel regression analysis.