Energy and Buildings 102 (2015) 163–171

Contents lists available at ScienceDirect

Energy and Buildings

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

Cross comparison of empirical and simulated models for calculating residential electricity consumption

Pamela Torres*, Michael Blackhurst, Nour Bouhou

The University of Texas at Austin, Civil Architectural and Environmental Engineering, USA

ARTICLE INFO

Article history:
Received 1 September 2014
Received in revised form
29 November 2014
Accepted 10 May 2015
Available online 21 May 2015

Keywords:
Residential energy model
Residential air conditioning demands
Multiple linear regression
Mixed effects modeling
RECS

ABSTRACT

The U.S. Energy Information Administration's Residential Energy Consumption Survey (RECS) provides electricity end-use estimates derived from disaggregating total household consumption. However, the accuracy of the EIA's disaggregation method is unclear, thus potentially producing erroneous end-use estimates and affecting respective decision support. In order to test the EIA's disaggregation methods, we perform parallel empirical analysis of residential cooling demands on two household samples located in Texas: an estimated sample (RECS) and a sample where air conditioning branch circuits are directly metered. Results indicate statistically significant differences between data sources, with the RECS sample possibly underestimating cooling demands. Results further indicate that models developed using observed end-use consumption are more robust, demonstrating statistically significant predictors that do not significantly correlate with end-use estimates from the RECS sample. We find that accounting for seasonality increases predicted R^2 values and reduces overestimation of statistical significance. Finally, the observed model indicates that window area and orientation are statistically significant predictors of cooling loads, which are parameters not surveyed in RECS. We then provide policy recommendations to improve end-use estimates, including integrating select utility-funded audit data into disaggregation methods.

1. Introduction

Many local, state and regional agencies have historically relied on simulated residential energy end-use consumption data for decision support, particularly for demand-side management [1–6]. Perhaps the most utilized end-use consumption estimates are published in the Residential Energy Consumption Survey (RECS), which the U.S. Energy Information Administration (EIA) has provided since 1979.

RECS is a nationwide cross-sectional survey that uses multiple sources to simulate and generate non-linear statistical models that estimate costs and usage for heating, cooling, appliances and other end uses [7]. EIA conducts in-person interviews to collect energy characteristics on housing unit (including dimension measurements), use patterns, and household demographics. EIA uses geographic indicators from other government agencies, such as cooling and heating degree-days (CDDs, HDDs), to complete the characteristics profile of each household.

Follow-up surveys with energy suppliers are conducted to obtain more accurate total annual energy consumption and expenditures. EIA [7] states that they received complete or partially complete data for 90% of the households surveyed and 90% participation of the energy suppliers. Missing and inconsistent data are edited using a variety of statistical adjustments. In brief, business end-uses are removed and total annual consumption is estimated where records are incomplete. End-use regression models and previous RECS publications are also used to quality assure and complete household records. Additionally, base-sampling weights for each household are calculated in order to scale sampled household estimates to national values from the RECS sample. Although energy data are collected on a monthly basis, RECS produces only annualized consumption and expenditures estimates for each reference year.

Consumption and expenditures for specific end-uses must be estimated by disaggregating total household consumption and expenditure estimates. The EIA [8] uses five separate and independently fitted nonlinear regression models—one for each of

Abbreviations: CCSP, U.S. Climate Change Science Program; EIA, Energy Information Administration; ICC, Intraclass Correlation Coefficient; ME, mixed-effects model; MLR, multiple linear regression; PS, Pecan Street Research Institute; RECS, Residential Energy Consumption Survey.

* Corresponding author at: The University of Texas at Austin Civil, Architectural and Environmental Engineering, 301 E. Dean Keaton, Stop 1752, Austin, TX 78712, USA. Tel.: +1 571 294 4621.
E-mail address: Pjt19@utexas.edu (P. Torres).

http://dx.doi.org/10.1016/j.enbuild.2015.05.015
0378-7788/

the main fuels (electricity, natural gas, propane, fuel oil, and kerosene)—for end-use disaggregation. Most of the independent variables are housing unit characteristics: physical home characteristics such as age and size of home; demographic information such as number of household members and income bracket; and equipment-related characteristics such as quantity or size. For example, total electricity use is an additive model of nonlinear models of space heating, air conditioning, refrigerator and other such appliances. The nonlinear models of individual end-uses (e.g., space heating or air conditioning) may include additional exogenous independent variables, such as cooling-degree days. Missing measurement data (i.e., non-energy data) is imputed using a statistical process called hot-deck imputation, where missing data is randomly assigned to the same data value as a similar household.

Final end-use estimations are based on an iterative procedure that uses nonlinear regression model techniques (SAS proc nlin with the Marquardt option). The estimation process uses end-use estimates from prior RECS publications, increasing or decreasing previous estimates based upon new information. Some adjustments are made using EIA expert judgment. For quality control, the summation of estimated end-uses may be further adjusted based upon the observed total consumption.

There are many examples of decision support and research using RECS. Recent examples include highlighting regional demand differences influenced by space conditioning balance points [9,10] and econometric analysis of residential energy consumption [11–18]. The U.S. Climate Change Science Program (CCSP) [19] recently used RECS data to estimate the future energy consumption for the buildings sector. Similarly, the U.S. Department of Energy has also produced multiple documents referencing studies reliant on RECS data. The National Action Plan for Energy Efficiency Leadership Group recommends RECS as a data source for demand-side analyses [20]. Several respective applications include residential energy efficiency simulations [21] and saving potentials due to smart metering [22].

Recently, however, some studies reveal discrepancies in the RECS data. For example, Kaza [23] shows that the 2005 RECS space-conditioning end-uses account for almost half of the energy consumption, while his analysis shows space-conditioning accounts for only 37% of the energy use. Kaza attributes this overestimation to the sampling weights used by the EIA that are not reflective of variability and lacks sample stratification method. More recently, Ewing and Rong [24] and Tso and Guan [25] performed multilevel regression analysis to capture the natural hierarchical structure of RECS, concluding that despite their more sophisticated and encompassing model, RECS fails to collect enough information to fully describe households' behaviors. Tso and Guan highlight inconsistencies with their results (e.g., less drafty homes consume more energy than more drafty homes), and suggest that increased precision of models requires more specific information of householders' end-use behaviors.

With emerging sources of high-resolution empirical data describing energy consumption and respective predictors, it is increasingly possible to explore, compare and validate the predictive power of RECS. Such analyses are important to identify any inappropriate assumptions supporting the RECS data that may consequently affect downstream decision-making processes. The purpose of our analysis is to develop empirical models of residential air conditioning demands and apply these models to both the RECS database (estimated) and a separate high-resolution dataset (observed) to develop insights into potential opportunities to improve RECS and similarly derived energy use estimates.

2. Methods

2.1. Data description

Observed electricity consumption data were obtained from the Pecan Street Research Institute (PS). PS is a 501(3)c non-profit that has partnered with The University of Texas at Austin and industry leaders to advance understanding of resource consumption in homes. The information is gathered through a series of home inspections, drawings, two household surveys, energy audits and metered energy consumption (electricity, water and natural gas) that encompass occupant characteristics (e.g., income, education, age, household size), physical descriptors of the home (e.g., floor area, home vintage, insulation) and behavioral and technology penetration choices (e.g., number of different energy efficient technologies, frequency of use of different technologies, indoor temperature settings). Though PS currently has over 1000 participants located throughout Austin, Texas, only 126 homes have completed comprehensive audits, which do not include demographic information such as income, education, household members, etc. Annual consumption for PS sample reflects the period between May 1, 2013 and April 30, 2014, which was selected to maximize the PS sample size with a complete annual record.

The 2009 RECS provides complete data for 12,083 households. To have comparable data sets, the RECS data was filtered to only provide urban, single-family detached homes in Texas, reducing the original 12,083 homes to 464 homes. Tso and Guan [25] and Kaza [23] both suggest models beyond multiple linear regression analyses, due to geographical factors that may influence consumption behaviors, i.e., households that share common social standards and a common environment may have similar consumption patterns despite different household features. Tso and Guan [25] state that most sampling designs of RECS (that obtain population represented samples) adopt multistage stratified sampling schemes, creating a multilevel structure where observations are not independently and identically distributed. The models specified by PS data, though obtained by multiple linear regression analyses, allow for natural clustering addressed by Tso and Guan because homes are geographically similar. However, in order to compare values across different models, the models specified by RECS data were also obtained using multiple linear regression analyses. The multilevel structural issue is addressed by matching 126 RECS homes to the 126 PS homes using one-to-one propensity score matching [26–30] based on total electricity consumption, cooling-degree days (CDD) and floor area.

While the PS data provides continuous insulation *R*-values, the RECS data provide a qualitative assessment of overall home insulation (poor, fair, good, adequate and excellent). Both RECS and PS insulation values were transformed into binary values, where homes are either adequately insulated (input = 1) or not (input = 0). The critical value for the continuous values of the PS data was calculated to be *R*-values above 17 [31].

Similar to Steemer and Yun [17], the distribution of each data set was checked for normality. The PS air conditioning electricity consumption is slightly positively skewed (0.42) and has greater than normal kurtosis (3.74). Nevertheless, a transformation was not pursued because doing so increases skewness and kurtosis significantly (minimum skewness of −1.77 and kurtosis of 6.31). No significant outliers were identified.

Though traditional analyses include income and education, energy audit data available for PS homes does not provide that information. Nevertheless, the exclusion of these indicators were not of great concern due to recent studies that consistently show the lack of statistical significance of income and education as directly impacting energy consumption [18,23,32].