The demand for household-level energy services: Estimating price elasticities using statistical learning

J. Wesley Burnett* and L. Lynne Kiesling†

May 3, 2021

Abstract

This paper presents a novel approach to estimate short-run price elasticities and economies of scale for household energy demand, while controlling for a high-dimensional dataset of housing, socio-economic, and behavioral characteristics. We explore household energy consumption in the context of recent national- and state-level trends in residential energy prices and changes to energy efficiency policies. To estimate elasticities across time, we use various years of data from the U.S. Energy Information Administration's ongoing Residential Energy Consumption Survey. Given the broad set of data offered by this survey, we use a statistical learning approach to analyze the most salient determinants of energy consumption and compare that approach to linear regression methods. The statistical learning estimator provides superior out-of-sample predictions of energy price elasticities relative to a reduced-form, linear regression approach. Moreover, we find increasing economies of scale in electricity but not natural gas expenditures. Finally, we discuss how statistical learning can be used for energy policy.

Keywords: LASSO estimation; Household energy demand; Price elasticity; Machine learning; Statistical learning.

*Research Economist, Economic Research Service, U.S. Department of Agriculture, 805 Pennsylvania Avenue, Kansas City, MO, 64105, wesley.burnettjw@usda.gov

†Visiting Professor of Engineering and Public Policy and Co-Director of the Institute for Regulatory Law and Economics, Carnegie Mellon University, 129 Baker Hall, Pittsburgh, PA, 15213, 412-268-2670, lynnekiesling@cmu.edu

Highlights

- Statistical learning estimator (LASSO) is applied to household energy demand.
- LASSO estimator better predicts price elasticities than linear regression models.
- Study finds energy price elasticity estimates that are consistent with theory.
- Estimates imply increasing economies of scale in electricity expenditures.
- Discusses how statistical learning can be applied to energy policy analysis.

1 Introduction

Residential energy consumption comprises almost one-third of total energy consumption in the United States (U.S. Energy Information Administration, 2020a). Average household energy usage in every region of the U.S. has declined steadily over the past decades. According to the U.S. Energy Information Administration (2020d), energy consumption per household has decreased by about 48 percent since 1980. During that same period, the average interior size of housing increased by 50 percent (U.S. Census Bureau, 2020a). These observations beg the question: How has household energy consumption continued to decline decade-over-decade despite rising square footage in new housing? Is the decline due to changes in behavior, household composition, technological innovation, government policies, or all of the above?

To answer these questions, we seek to understand how households make energy consumption decisions within a technological and institutional context. According to the International Energy Agency (2019), technological improvements to energy efficiency has led to approximately $145 (U.S. dollars or USD) in savings per household (or about an eleven percent reduction in annual energy costs) in the U.S. since 2000. These household saving estimates are for non-transportation related expenses, and the International Energy Agency (2019) argues that households save an additional $27 due to energy efficiency gains in transportation. As an example, the adoption of energy efficient appliances has led to a gradual reduction in household energy use over time (Lovins, 1988; Young, 2008; de la Rue du Can *et al.* , 2014).

These changes depend, among other factors, on how sensitive consumers are to energy price changes. Price elasticity of demand is an important variable in residential energy economics models. Both the theory and practice of regulated rate design incorporate price elasticity and its differences across residential, commercial, and industrial customers. In wholesale power market regulatory policy and market design, price elasticity of demand has important implications for price volatility and the incidence of price spikes. And, as we discuss in this paper, price elasticity estimates have been a focus of research on residential energy consumption and trends in energy use. Rate designs and market designs that rely on poor estimates of price elasticity can yield unintended consequences and affect the distribution of costs across different types of customers, an issue that is particularly contentious for the residential customers that are the subject of this analysis.

The price elasticity of demand indicates the degree of substitutability that consumers perceive in a market. Understanding such substitutability will become increasingly important as digitization and distributed energy resources like rooftop solar PV and electric vehicles expand in the future, and residential consumers have more pricing, self-supply, and storage alternatives.

Thus, estimating price elasticity is an essential element of understanding energy systems. Price elasticity estimates have been a focus of energy economics for the past four decades (Dubin & McFadden, 1984; Baker & Blundell, 1991; Ryan & Plourde, 2009), and have progressively incorporated methodological advances in estimation techniques. In this paper, we apply novel statistical (or machine) learning techniques to estimate residential price elasticities, building on and extending existing models.

Since the 1980s, a large and varied literature has emerged on the determinants of household energy consumption and the econometric methods used to estimate this demand. Our paper adds to this literature by introducing a data-driven method to estimate the determinants of household demand. This paper's primary contribution is to use statistical learning techniques to shrink the number of potential covariates, which optimizes the specification for household demand.[1] Machine learning is a branch of artificial intelligence that gives computers the ability to learn high dimensional patterns from raw data (Jordan & Mitchell, 2015).[2] Using the U.S. Energy Information Administration's Residential Energy Consumption Survey (RECS), we estimated residential energy short-run price elasticities conditional on various physical housing characteristics and the socio-economic makeup of the household members.

The justification for this new contribution to the large literature on household energy demand is *model selection*. Many past studies are based on least squares analyses with ad hoc choices for co-

[1]According to some experts, machine learning and statistical learning are two separate and distinct frameworks or areas of research (Stewart, 24 March 2019). Without wading into this debate, we offer the two terms interchangeably to mean statistical predictive models based on algorithm used within the machine learning literature.

[2]Artificial intelligence is a broad interdisciplinary field that seeks to make intelligent machines or intelligent computer programs.

variates. In reality, though, we do not know the underlying true model, nor which determinants are the most important for inclusion in a model. Two common problems in linear regression modeling are overfitting and omitted variable bias. Overfitting occurs when the researcher includes too many regressors, resulting in a good in-sample fit to the data (e.g., a large estimated R-squared value), but poor out-of-sample prediction. Conversely, if one includes too few regressors, then the model may suffer from omitted variable bias.

Model selection can be challenging when the data are high-dimensional; that is, the number of regressors is close to or larger than the number of observations. The high-dimensional data problem can be addressed through regularization (or shrinkage), in which the non-essential predictors are shrunken toward zero (relative to least squares estimates). This shrinkage method helps to reduce variance (in the statistical learning sense of the term), and such an approach can also be used to perform variable selection (James *et al.* , 2013).

Having an unknown true model *ex ante* is not uncommon, particularly when modeling household energy demand. Regularization methods, discussed more specifically below, has been demonstrated to outperform least squares in terms of out-of-sample prediction. Belloni *et al.* (2014a) illustrate how statistical learning can be used for estimation in applied econometric analysis. We follow the least absolute shrinkage and selection operator (LASSO) approach by using a double-selection methodology to estimate household energy demand.

More specifically, we estimate price elasticities of demand (as well as economies of scale in expenditures) for household energy consumption for survey years in 2001, 2005, 2009, and 2015.[3] This study's primary objective is not necessarily identifying the causal effect of energy prices on consumption, but rather comparing ordinary least square estimates to statistical learning estimates. Gelman & Imbens (2013) argue that a causal approach focuses on the "effects on causes," whereas a non-causal approach (as is the case in our study) focuses on the "causes of effects." We follow the latter approach, in the context of traditional statistics, to conduct model checking and hypothesis generation.

[3]The RECS survey is not longitudinal, and survey data is not available for the intervening years between these dates.

We find that the statistical learning approach improves the out-of-sample prediction (i.e., reduces the mean-square errors) and improves the within-sample fit to the data (i.e., produces larger R-squared estimates) of total household energy demand relative to least squares estimates. Based on our findings, we discuss how these techniques can be used to improve policy analysis.

2 Estimating Residential Energy Demand

Energy use in homes provides heating, cooling, lighting, refrigeration, and other services that have become essential in the modern economy. An average U.S. household spends 3.3 percent of their household budget on energy services (U.S. Bureau of Labor Statistics, 2020a). Understanding household energy demand has been the focus of a robust microeconometric literature examining the factors determining energy demand and the price and income elasticities of demand (both short run and long run) as behavior, technology, and policies change through time.

As an input into many residential consumption activities, energy consumption and expenditures are a function of demographic characteristics of household members, household income, physical characteristics of the home, and the nature and composition of energy-using devices in the home. These factors can all influence how responsive (along both the intensive and extensive margin) a household is to changes in energy prices.

Household energy demand is thus a derived input demand that also depends on the housing stock and durable consumer assets in the home. This complicated nature of household energy demand has generated considerable work on accurate econometric specifications to estimate models of household energy demand and the associated price elasticities. Bohi & Zimmerman (1984) and Berndt (1991) provided early surveys of the variety of estimation techniques, finding that estimates of demand behavior vary considerably depending on the estimation methods and the data sample analyzed. Fouquet (2014) analyzed 200 years of British consumption data to estimate long-run price and income elasticities of energy demand. His reduced-form econometric analysis also provided a valuable survey of the extensive literature on energy consumption.

Dubin & McFadden (1984) developed an econometric model of residential energy consumption to analyze household energy consumption in the U.K. from 1972 to 1988. Their model accounts for the joint determination of demand for energy and demand for appliances in estimating unbiased long-run elasticity parameters. Using pooled household-level annual data from the Family Expenditure Survey, Baker and Blundell (1991) find that energy demand varies considerably across households, and the main determinants of that demand were energy-using appliance ownership, renting or owning the home, and age of head of household. They also find that the demand for both electricity and gas were inelastic, but gas demand was more inelastic while electricity demand elasticity was close to unit elastic. Further, they found income elasticity estimates that were less than unity, suggesting that both electricity and gas are essential household consumption goods. Ryan & Plourde (2009) highlight the fact that most of the past household energy demand studies rely on reduced-form models, which partially explains the differences in findings over time.[4]

Huebner *et al.* (2015) used cross-sectional data on UK residential energy consumption to explore the role of behavioral factors. They found that physical building characteristics were the most important factors in energy consumption, and other significant factors were household size, length of heating season, and household member beliefs about climate change. Their analysis is closest to ours because of their use of the LASSO method for minimizing confounding covariates in their high-dimensional data.

Our analysis contributes to the literature on household energy consumption by applying statistical learning (SL) techniques. Athey & Imbens (2017) provide an overview of SL techniques that can be useful in empirical policy evaluations. SL techniques for economic data analysis "have been particularly successful in big data settings, where we observe information on a large number of units, many pieces of information on each unit, or both, and often outside the simple setting with a single cross-section of units" (Athey & Imbens, 2019, p. 686). When combined with econometric modeling, SL techniques can improve the efficiency of least-squares/maximum likelihood estimators. In the context of a question like household energy consumption, with big data and a

[4]Ros (2017) estimated price and income elasticities of demand using utility-level panel data. His reduced-form and structural estimations of residential, commercial, and industrial demand were consistent with previous estimates using aggregated data.

sparse theoretical model, SL techniques can reduce overfitting of the model that could result from inclusion of too many covariates.

3 Data

The household-level data used in this analysis come from the Energy Information Administration (EIA) Residential Energy Consumption Survey (RECS). RECS is roughly a quadrennial survey with an extensive list of questions including home size, number and type of appliances and their energy use, number of residents, and other demographic information. The EIA has conducted the RECS survey since 1978.

3.1 RECS Survey Administration

The EIA administers the RECS as a two-phase survey. Phase 1 is a multi-stage survey of a nationally representative sample of households, conducted through online, mail, and interview methods to collect demographic information and energy usage patterns. They apply a complex process of stratification by census region to ensure a representative cross-section sample. Phase 2 is an energy supplier survey to gather billing data on usage and expenditure for the sampled households.

After gathering these data, EIA analysts then use the survey data to construct models of disaggregated energy consumption and expenditures by end use for each household. The most recent surveys occurred in 2001 (4,882 households), 2005 (4,382), 2009 (12,083), and 2015 (5,688).[5] The exact questions vary across years, but a large number of consumption and demographic questions have been included in all four surveys. The online appendix provides a list of the questions and variables that are common across the four most recent surveys. The RECS is not a longitudinal survey, so it does not contain the same households in subsequent years. This survey design will have implications for our choice of econometric estimation methods.

The methodology for the 2015 survey differs from previous years. In past years, the EIA used

[5]Sample sizes are reported on the EIA RECS website: RECS 2001 survey data, RECS 2005 survey data, RECS 2009 survey data, RECS 2015 survey data.

regression models to estimate household-level consumption and expenditure, and calibrated their models using a normalization procedure. In 2015 they changed the model estimation technique to an engineering-based approach that uses existing studies with physical data to model consumption and expenditures with improved expected accuracy. The calibration approach is now a minimum variance estimation.[6]

3.2 Descriptive Statistics

Table 1 reports mean and standard deviation for a common set of variables across all survey years between 2001 and 2015. The first variable in the table, type of home, is a categorical variable indicating the type of home in which the respondent resides (mobile home, single family home, multi-family home, small apartment building, or large apartment building). Based on the data across survey years, approximately 63 to 68 percent of respondents resided in single family homes. The next largest percentage resided in large apartment buildings. The mean heated interior space (total square feet) varied between 2,000 and 2,300 square feet over the entire sample. The U.S. Census Bureau (2020a) reports that the median size of a new single-family home in 2019 was about 2,300 square feet, so the RECS observations are generally consistent with the national average.

According to the results in Table 1, the mean level of household energy consumption (represented by total British Thermal Units or BTUs) increased from 2001 to 2005, but then declined in 2009 and 2015. These trends are consistent with the EIA's (2019c) findings that energy consumption per household has declined over the past two decades. The EIA contends these trends are due to improvement to building insulation and improved efficiencies in household appliances; Davis (2017) also argues that increased energy efficiency in lighting has contributed substantially to the decrease. The data also show a steady increase in the number of respondents who have identified using indoor air conditioning, making the recent decrease in mean energy consumption all the more striking.

In terms of socioeconomic demographics, the head of household's average age has remained at

[6] More information can be found at `https://www.eia.gov/consumption/residential/reports/2015/comparison/index.php`.

approximately 50 years across all of the survey years. Approximately fifty percent of respondents identified as being employed and about thirty percent reported owning their home. Also, the average number of household members across all survey years was approximately two-and-one-half persons.

Annual electricity expenditures constitute the largest reported energy costs across the survey years, followed by natural gas. Electricity costs, on average, ranged between \$960 and \$1,400 per household, and the reported average electricity expenditures increased for each subsequent year of the survey. Annual natural gas expenditures, on the other hand, ranged between an average \$330 and \$780. As noted above, the reported average expenditures for natural gas are highest in 2005 and lowest in 2015. The drop in expenditures is likely due to the shale gas revolution and the subsequent increase in supply of domestic natural gas. Otherwise, reported heating oil expenditures were fairly stable, with the exception of 2005, in which annual costs were nearly double that of the other reported years. This increase is arguably due to the increase in crude oil (heating oil is a derivative from the refining process of crude oil) prices up to the year 2005. The average (nominal) spot price of West Texas Intermediate reached \$56, which constituted an approximate 72 percent increase over the previous three-year average (U.S. Energy Information Administration, 2021a). [7]

3.2.1 RECS Data Preparation

The RECS offers a rich array of housing and household member characteristics to ascertain the demand for residential energy services. Nonetheless, one of the primary shortcomings of the survey (in its format over the past two decades) is that its lack of energy price data.

Despite this fact, one can calculate the implied unit energy prices by dividing the household's energy expenditures by the housing unit's total annual energy consumption for a particular resource. For example, the RECS survey reports a household's total electricity consumption in units of kilowatt hours (KWh) and its total annual electricity expenditures in dollars. We estimated the

[7] The average reported expenditures on propane and kerosene were very small. Annual expenditures on kerosene were no longer surveyed starting in 2015, and the reported number of households that consumed kerosene in 2009 were so small that we omitted the observations from the study.

unit prices for natural gas, heating oil, propane, and kerosene in a similar fashion. Theses implied energy prices are arguably endogenous to the household's energy consumption behavior. However, the primary objective of this study is not necessarily to appeal to causal inference but rather to predictive inference. To ensure the accuracy of the implied prices, we compared the estimates to state-level residential electricity prices reported by U.S. Energy Information Administration's (2018) Annual Electric Power Industry report (EIA-861).

3.3 Trends in the Prices and Consumption of Electricity and Natural Gas

To analyze the changes in the estimated price elasticity of demand for electricity and natural gas, we explored changes to national and state residential demand. Figure 1 illustrates national average, per-capita electricity and natural gas use for the years 2000 to 2018 (the last year of observation). The points in the figure represent yearly calculated averages in megawatt hours (MWhs) and thousands of cubic feet (Mcf), respectively. We fit a second-order polynomial trend line to demonstrate the change in per-capita consumption. Electricity usage increased between 2000 and 2008, which demarcates the beginning of the Great Recession. Since that peak, consumption has declined steadily through 2018. Natural gas demand, on the other hand, declined over the entire period of observation.

Figure 2 illustrates average, state-level usage comparing the decade 2000-2009 and the subsequent decade 2010-2018. Average consumption for the 2010s is on the vertical axis, and average consumption for the 2000s is on the horizontal axis. The straight line the 45 degree angle within the graph. Any points (states) that lie above (below) the 45 degree line show an increase (decrease) in per-capita consumption relative to the preceding decade. We labeled a few of the states for expository purposes. Of the 50 states (and the District of Columbia), 34 experienced a decrease and 17 experienced an increase in electricity consumption.[8] Only six states experienced an increase in the average consumption of natural gas.[9]

[8]States that experienced an increase in average consumption were Arkansas, Iowa, Kansas, Louisiana, Maine, Michigan, Missouri, Mississippi, Montana, North Dakota, Nebraska, New Mexico, Oklahoma, South Dakota, Utah, West Virginia, and Wyoming.

[9]States than experienced an increase in average consumption of natural gas were Connecticut, Idaho, Louisiana, New

What has driven the decreases in residential electricity and natural gas consumption? According to Davis (2017), the decline in electricity usage is due primarily to the adoption of energy-efficient technologies, such as LED light bulbs. By 2017, the average residence was consuming 10.4 megawatt-hours (MWhs) of electricity annually, which is down from the average of 11.5 MWhs consumed in 2010 (McGinty, October 11 2019). As a result, residential electricity consumption dropped from 129.7 million MWhs in 2015 to 95.5 million MWhs in 2017 (Ayesh, October 11 2019). Moreover, Davis (2017) argues that many households have adopted energy-efficient appliances, including refrigerators, water heaters, air conditioning, and dishwashers, among other. These improved efficiencies, combined with better building insulation and population migration to regions with lower heating needs has led to an overall decrease in energy use per household (including natural gas consumption) (U.S. Energy Information Administration, 2020d). Nadel *et al.* (2015) claim that energy efficiency is the primary contributor to a 37 percent decline in energy use per household.

Other potential causes of declining energy use exist, including price and income effects. The decrease in household energy demand is unlikely to arise from income effects, as energy is a normal good and U.S. household income increased by about 13 percent over the period of 2000-2018 (Economic Policy Institute, 2019). Furthermore, our income elasticity estimates in Table 2 are positive and statistically significantly different from zero across all survey years, suggesting that income effects were not a factor in decreasing demand.

To address the question of potential price drivers, we explored changes to residential electricity and natural gas prices. Figure 3 displays average (national) residential electricity and natural gas prices in both nominal and real terms. The nominal residential price of electricity increased between 2000 and 2008, flattened between 2008 and 2012, and then started to climb after 2012. In comparison, the real price of electricity remained essentially flat over this entire period. Nominal (residential) natural gas prices, on the other hand, increased until about 2007 and then declined thereafter.[10] The real price of natural gas declined over this period.

York, Rhode Island, and Vermont.

[10]The decline of the nominal price of natural gas was arguably caused by the large increase of unconventional (hori-

Figure 4 depicts (nominal) state-average prices in the 2000s (horizontal axis) and in the 2010s (vertical axis). As with the previous figure, any points (states) that lie above the 45 degree line represent an increase in average prices over the previous decade. All state-level (nominal) average electricity prices increased over the previous decade. During this decade, Hawaii and Alaska had the highest average residential electricity prices at approximately \$0.32 and \$0.19 per Kilowatt-hour (KWh). For residential natural gas prices, only 18 states had increased prices, whereas 33 states (including the District of Columbia) had lower prices in the 2010s.

To explore further the determinants of household energy demand, we present the results for the statistical learning, post-model-selection approach in the next section.

4 Methodological Approach

Each year's RECS reflects survey results of different households (i.e., not repeated surveys of the same households), so the RECS is not a longitudinal survey. Hence, pooling the different years of household-level data into a single panel (such as, pooled ordinary least squares) would generate biased and potentially inconsistent estimates. Even if it were appropriate, pooling the data and then using either a random effects or a fixed effects model would assume that the slope coefficients were the same across the households.[11] Independent cross-section estimations, on the other hand, assumes that these households, while not identical, do not have common characteristics that would account for their energy consumption decisions across time. Based on these data limitations, our study focuses on the heterogeneity of demand elasticities across the separate survey years.

Reflecting the fact that the real relationship is likely to be some combination of the estimates using the pooled and the cross-sectional data, an accurate parameter estimate would be a weighted average of estimates from the two models. An econometric method that addresses this challenge is the shrinkage estimator. The logic of a shrinkage estimator is to allow for heterogeneity within a maximum likelihood or generalized least squares framework, while at the same time generating

zontal drilling and fracturing) natural gas production in the U.S. (Hausman & Kellog, 2015).

[11]Note that this issue applies to the Baker & Blundell (1991) analysis using pooled data from the UK FES, which is also not a longitudinal survey.

asymptotically consistent and unbiased parameter estimates.

Maddala *et al.* (1997) used a variety of shrinkage estimator specifications to estimate short-run and long-run elasticities of residential electricity and natural gas demand. They used a structural demand model different shrinkage estimator methods to estimate the model.[12] They found that that shrinkage estimators yielded more realistic parameter estimates than either the pooled or cross-sectional models. Our analysis applies the logic from Maddala *et al.* (1997) to a reduced-form estimation of short-run price and income elasticities as discussed in Ryan & Plourde (2009).

4.1 The LASSO Estimator

The least absolute sum of squares operator (LASSO) estimator finds a solution for the linear equation $y = \sum_{i=1}^{p} \beta_i \, x_i + \varepsilon$, where *p* denotes the (high-dimensional) number of potential regressors. The lasso method minimizes the prediction error subject to a constraint. One of the more common estimation methods of the LASSO is to minimize

$$\frac{1}{2\,N}\,(\mathbf{y} - \mathbf{X}\,\boldsymbol{\beta}')'(\mathbf{y} - \mathbf{X}\,\boldsymbol{\beta}') + \lambda \sum_{i=1}^{p} |\beta_i|. \tag{1}$$

The LASSO estimator is a form of penalized regression in which the estimator's algorithm minimizes the sum of squared errors while simultaneously constraining some parameter estimates to zero, enabling the shrinkage in the covariate set. The amount of shrinkage is a function of the chosen penalization parameter λ. In essence, this technique prioritizes inclusion of covariates that have relatively more explanatory power than others, and the penalization parameter determines the cutoff point. As discussed in Maddala *et al.* (1997), the logic of the shrinkage procedure is similar to that of weighted least squares, in the sense of achieving consistent estimates by correcting for endogenous sampling (Solon *et al.* , 2015).

More specifically, the first term, $(\mathbf{y} - \mathbf{X}\,\boldsymbol{\beta}')'(\mathbf{y} - \mathbf{X}\,\boldsymbol{\beta}')$, in equation (1) is the in-sample prediction error, which is the same value that least squares estimation minimizes. The second

[12] The three shrinkage estimators they used were Stein (classic), Bayesian two-step, and Bayesian iterative. The choice of estimator depends on the structure of the data.

term, $\lambda \sum_{i=1}^{p} |\beta_i|$, is the penalty that increases the value the more complex the model (StataCorp, 2019).[13]

The LASSO estimator proceeds by choosing a solution to the minimization problem based on the out-of-sample prediction error (StataCorp, 2019). If the tuning parameter λ is equal to zero, then the estimation procedure reduces to least squares. Three different methods exist for estimating λ: plugin, cross- validation ("CV"), and adaptive ("Adapt"). Cross-validation selects the tuning parameter so to minimize the out-of-sample prediction error. The adaptive method performs multiple lasso estimates, based on cross-validation, and variables with zero coefficients are eliminated and the remaining variables are given penalty weights designed to weight small coefficients to zero (StataCorp, 2019). The plug-in method creates a tuning parameter that is typically larger than the CV method, and therefore, there are fewer regressors in the final model (StataCorp, 2019).[14]

4.2 Applying LASSO to the RECS Data

The RECS data include a large number of variables and a somewhat large number of households per year, and the data are high-dimensional because of the large number of variables relative to the sample size. Analyzing the determinants of household energy consumption and short-run elasticities of demand involves estimating a large number of parameters, in a context with many potential covariates and no strong economic theory to guide in regressor selection. In other words, estimating with many parameters of interest runs the risk of overfitting the model, which potential creates a degrees of freedom challenge.

The LASSO estimator is one of the main approaches to regularization in statistical learning. Regularization, or reducing the dimensions in high-dimension data, contributes to proper model selection (Belloni *et al.* , 2014b), and thus aids in the ability to draw conclusions using high-dimensional data. The conventional, ad hoc method of selecting covariates for parametric and nonparametric approaches is an example of regularization, but one that relies solely on the re-

[13] Another variant of the above equation includes regression weights within the constraint.

[14] The interested reader can learn more about this technique from Belloni *et al.* (2012), Belloni *et al.* (2014a), Belloni *et al.* (2014b), and Athey & Imbens (2017), among others.

searcher's ability to determine which covariates to include and how (Belloni *et al.* , 2014a). Using high-dimensional data to analyze empirical phenomena could take advantage of the data-driven methods offered by regularization.

Our particular empirical approach follows what Belloni *et al.* (2014b) describe as the "post-double-selection" methodology, which uses the LASSO estimator to select control variables. To better understand this methodology, consider the model

$$y_i = d_i^{'}\alpha + x_i^{'}\beta + \varepsilon_i, \tag{2}$$

where y_i is again the outcome variable, d_i denote a vector of the covariates of interest, x_i is a vector of controls, and ε_i is a random noise term. The term x_i constitutes a high-dimensional set of potential explanatory variables. As identified by Belloni *et al.* (2014b), the "right" set of controls is not known *a priori*. Choosing too few regressors can lead to omitted variable bias, while choosing too many regressors can result in overfitting.

The post-double-selection methodology uses the LASSO twice. In the first step, a LASSO regression is estimated for y_i on x_i. In the second step, a LASSO regression is estimated for d_i on x_i. Regularization (or shrinkage) reduces most of the coefficients to zero. The final step of the estimator procedure uses the union of controls selected in steps one and two to estimate a final least squares regression of y_i on d_i. This procedure is used to construct orthogonalized estimates of the dependent variable on the exogenous (or near exogenous) variables of interest. Chernozhukov *et al.* (2015) offer a similar approach to the post-double-selection methodology that includes a vector of instrumental variables.

5 Results

5.1 Traditional least squares approach with model uncertainty

We estimated each cross-sectional survey year via ordinary least squares (OLS). Table 2 presents these first estimates, based on an ad hoc, reduced-form specification of household energy demand including variables used in previous studies. Several conditional variables were estimated in addition to what is presented in the table, but those covariates are omitted here for the sake of space. The entire regression estimates can be found in the online appendix.

The short-run price elasticity for electricity is in the first row labeled "Electricity." The estimate for the 2005 survey year implies that a ten percent increase in electricity prices led, on average, about a two percent diminution of electricity consumption. The estimated elasticities imply that the demand for residential electricity is highly inelastic, meaning that consumers were not responsive to price changes. These estimated elasticities are generally consistent with many previous findings (Taylor, 1975; Bohi & Zimmerman, 1984; Maddala *et al.* , 1997; Bernstein & Griffin, 2005; Paul *et al.* , 2009; Alberini & Filippini, 2011; Deryugina *et al.* , 2017). As an example, an Electric Power Research Institute (2008) report found that average estimated short-run elasticities ranged between -0.2 and -0.6.

The price elasticity for natural gas is provided in the row labeled "Natural gas" of Table 2. The estimated elasticity for 2001 implies that a ten percent increase (decrease) in natural gas prices led to an almost four percent decline (increase) in energy consumption. The estimated coefficients on natural gas prices also imply a relatively inelastic demand, alhough our estimated elasticities are slightly larger in magnitude (in absolute terms) than the findings of Auffhammer & Rubin (2018), who found estimates that ranged between -0.17 and -0.23.

The estimated price elasticities for heating oil, propane, and kerosene are offered in the subsequent rows of Table 2. The number of survey responses for these fuels were far fewer than the reports of electricity and natural gas use. Based on the small number of responses, we do not know if the household actually does not consume these resources or if their use were missing (not reported). To circumvent this potential problem we used the dummy variable adjustment method outlined in Allison (2010).

The short-run price elasticity for heating oil in the 2009 survey suggests that a ten percent increase in prices led to an approximate nine percent decline in total household energy consumption, which is the only significant estimate among the survey years observed.[15] The short-run price elasticity of demand for propane implies that a ten percent increase in propane prices led to about a six percent decrease in propane consumption. The kerosene price elasticities were omitted for the 2005 and 2009 survey years as there were too few responses, and the EIA aggregated heating oil and kerosene prices into the same category in the 2015 survey year.

The elasticity estimate of square footage (a measure of a house's cooled and heated interior space transformed to natural logs) is positive and significant across all survey years in Table 2. The estimated coefficient for the 2001 survey suggests that a ten percent increase square footage would generate about a one percent increase in energy consumption. Moreover, the estimated coefficient on the number of persons residing in the household ("No. household members") suggests that another person increases total energy consumption by about seven percent. Finally, the age of the head-of-household ("Householder age") implies a small positive, but mostly insignificant, effect on household energy consumption.

Following Ironmonger *et al.* (1995), we estimated the economies of scale in energy expenditures of each household by transforming the count variable of household members, within the RECS survey, to a factor variable (i.e., we created a binary variable for the reported number within each household). The estimates are based on a linear regression of total annual energy expenditures on the member number factor variable and the total square footage within the home. Table 3 provides the scale elasticity estimates for electricity expenditures, while Table 4 presents estimates for natural gas expenditures.

The marginal effects for the reported number of household members are reported in Panel A of Table 3. To determine the economies to scale we contrasted the marginal effects of different household sizes. More specifically, we estimated the differences in the numerical derivatives based on the number of household members (from the prediction function of electricity expenditures). The coefficient for the marginal effect of two persons (members), for the 2001 survey year, implies that

[15] The increase in the absolute magnitude of the elasticity estimate in 2009 is arguably due to the run-up of residential heating oil prices in 2008, which coincided with the increase in crude oil prices. The prices rose by approximately 72 percent in 2008 (over the previous five-year average) (U.S. Energy Information Administration, 2021a).

a two-person household spends about $200, on average, compared to the sample mean. Intuitively, the estimated marginal effect increases as the number of persons within the household gets larger, and the estimates are consistent across the survey years.

The marginal effects do not reveal whether smaller households pay more or less for electricity relative to larger households. To estimate the differences in expenditures, based on the number of persons within the household, we contrasted the predicted expenditures between each subsequent household size. The contrasted expenditures are offered in Panel B of Table 3. These findings provide evidence of increasing economies to scale for electricity expenditures within the home. For example, the first estimated coefficient, under the 2001 survey year, implies that a one-person household spends about $200 less than a two-person household, and a two-person household spends about $113 less than a three-person household. The predicted differences in household expenditures are statistically significant for up to five members, after which the estimates are no longer different from zero. This result provides limited evidence of a threshold effect, in which additional persons within the household, beyond five people, does not create any significant differences in electricity expenditures. As a robustness check, we also contrasted predicted expenditures between each category and the overall sample mean of expenditures (not provided, but available upon request), and the findings were qualitatively similar.

In addition to electricity, we also examined scale economies with annual natural gas expenditures, which are provided in Table 4. As before, the marginal effects of household size is offered in Panel A, whereas the difference in expenditures is provided in Panel B. The marginal effects estimates for each increasing household size category are mostly statistically significant. Unlike the previous table, though, there is little to no evidence of economies of scale in natural gas expenditures.

The initial results in Tables 2 through 4 offer similar elasticity estimates that are consistent with economic theory. However, all of the linear regressions are based on ad hoc choice of regressors due to a lack of knowledge of the true underlying model.

5.2 LASSO Estimation

Given the model uncertainty within the linear regression models, we apply a post-model-selection approach using the LASSO estimator.[16] To produce the estimates, we used the LASSO estimator for each separate survey year from 2001 to 2015, applied to the full set of potential explanatory variables within each survey year.

Table 5 presents price elasticity estimates based on the LASSO estimator. For the sake of comparison, we have included the linear regression estimates from Table 2 under column (1) of each survey year. The LASSO estimator provides a regularization method to reduce the dimensionality of the regressand set.[17] The process of shrinkage requires an estimate of the penalization or "tuning" parameter (i.e., the λ parameter outlined in Belloni *et al.* (2014a, p. 32) or Belloni *et al.* (2014b, p. 615)). As stated above, this parameter (along with regression weights if applicable) shrinks the non-informative explanatory variables towards zero. As a robustness check we estimated the tuning parameter(s) by all three candidate methods described in Section 4.

To estimate the models and compare the predicted values to the observed, we *a priori* split each year's sample randomly to define seventy-five percent of the observations as in-sample and twenty-five percent as out-of-sample. The double-selection linear model estimates were performed on the defined in-sample set of observations. We then estimated the mean-squared-error (MSE) and R^2 for each predictive model. The MSE and R^2 estimates were calculated by comparing the out-of-sample predicted values to the actual out-of-sample observations. The number of out-of-sample predicted values are listed in the rows labeled as "OS obs" in Table 5.[18] Better predictive models have lower estimated MSE values and larger estimated R^2 values. That is, the model offers good in-sample fit to the data as well as good out-of-sample prediction.

For the sake of space, we only compare the energy price elasticities; however, the full set of

[16]The estimation procedure is explained in Section 4.

[17]As part of this process, some of the potential regressors are omitted to improve the model's out-of-sample predictive abilities.

[18]The number of out-of-sample observations differ between methods as observations are dropped due to multicollinearity within the data.

estimation results are available in the paper's online repository.[19] In general, the estimated price elasticities for residential electricity, natural gas, and propane are statistically different from zero. The estimated price elasticities for heating oil, on the other hand, are only statistically significant for the 2009 survey year, reflecting the higher oil prices discussed earlier. Hence, the residential price of heating oil seems to have significantly affected household energy demand as reported in the 2009 survey year. The relative effect of these prices is implied by the magnitude (in absolute value) of the estimated coefficients, which is larger than the other energy price elasticities for that same survey year. In other words, households appear to have been relatively more responsive to price changes in heating oil than the other energy sources.

The magnitudes of the estimated coefficients on natural gas prices, across the survey years, correspond with the average price time series observed in Figure 3b. The magnitudes for the estimated elasticities in the 2009 survey year are larger than the other survey years, consistent with national average prices of natural gas reaching a peak of nearly $14 per thousand cubic feet in 2008. National average prices fell after 2009, reaching a price of approximately $10 per thousand cubic feet in 2015, and the magnitudes (in absolute value) of estimated price elasticities for natural gas are smaller for the 2015 survey year.

Similarly, the magnitudes of the estimated coefficients on electricity price elasticities are largest (in absolute value) for the 2015 survey year. As depicted in Figure 3a, the (nominal) national average prices of (residential) electricity reached nearly $0.13 per kilowatt hour in 2015, which is up from an average price of $0.07 per kilowatt hour in 2001. Thus, the estimated elasticities are much larger (absolute value) for the 2015 survey year compared to the 2001 survey year. Again, these findings imply that households were more responsive to the price changes in the later years of the survey relative to the price changes in the earlier years of the survey.

National average propane prices started at approximately $1.25 (nominal value) in 2000, reached a historical high at nearly $3 per gallon in 2011, then tapered off to a price just under $2 by 2020 (U.S. Energy Information Administration, 2021b). The estimated elasticities across the survey

[19] `https://github.com/lkkinetic/RECS-LASSO`

years seem to mimic the pricing behavior as the elasticities in the 2015 survey year (national average price was $2.87 per gallon in 2014) are consistently the largest (in absolute terms) relative to the previous survey years.

Lastly, the double-selection LASSO linear models provide better predictions for the out-of-sample observations as demonstrated by the smallest MSE estimates, and a better fit to the in-sample data as demonstrated by the R^2 estimates. Specifically, the cross-validated method of estimating the λ parameter performs better on both measures across all of the survey years.

The double-selection linear model estimates are based on the full set of variables within each of the survey years. An argument against this approach is that the survey questions differed across the years, with the 2009 survey offering the broadest set of questions and the largest number of observations. To address this potential shortcoming, we also estimated the double-selection linear models based on the common set of questions asked in each of the observed surveys. For the sake of exposition, those estimates are offered in the online appendix. Despite the smaller covariate set available within each survey year, the results are qualitatively similar to the estimates in Table 5. The cross-validated or adaptive method for estimating the LASSO offered better out-of-sample predictions and within-sample fit to the data.

6 Conclusion and Policy Implications

This analysis contributes to the burgeoning literature on statistical learning techniques used in modeling residential energy consumption. Our study is based on the U.S. EIA's ongoing Residential Energy Consumption Survey, which provides a high-dimensional set of data on the potential determinants of total household energy use. Past studies have largely relied on ad hoc, linear models due to model uncertainty and a lack of theory. Such approaches generally provided a good in-sample fit to energy consumption data, but lack out-of-sample predictions.

Using the least absolute shrinkage and selection operator or LASSO estimator, we showed that statistical learning can not only provide better in-sample fit, but also superior out-of-sample predictions (relative to learning modeling approaches) of household energy consumption. The use of statistical learning techniques can help stakeholders improve energy efficiency and ultimately reduce energy consumption. Combined with other state and national programs (such as energy efficiency and building codes), statistical learning techniques could potentially lead to substantial savings for U.S. households and the power sector in general.

Machine or statistical learning is being used increasing more often for the energy industry and policy. There are many potential applications of statistical learning in energy including predictive maintenance (risk management), predictive pricing, demand forecasting, trading strategies, data processing, and renewable energy forecasting, among others (Ghoddusi *et al.* , 2019). In energy policy, accurate estimation techniques for high-dimensional data can contribute to improved analyses for energy efficiency, weatherization, and low income assistance programs.

Specific examples of successful statistical learning applications within industry include DeepMind, Open Climate Fix, and the Open AI Energy Initiative. DeepMind partnered with Google to apply machine learning algorithms to 700 megawatts of wind power capacity in the central United States. According to Witherspoon & Fadrhonc (26 February 2019), DeepMind can predict wind power output 36 hours ahead of generation, and therefore, has increased the value of wind energy by about twenty percent. These better forecasts increase the value of renewable energy and reduce the requirements for backup energy from fossil fuels.

Open Climate Fix is a non-profit research and development lab that uses open-source satellite imagery data to forecast solar photovoltaic (PV) power production (Open Climate Fix, 2020). Additionally, this non-profit is using other forms of open-source data (including OpenStreetMap) to improve energy efficiency and offers maps for siting current PV installations.

Moreover, Shell, Baker Hughes, and Microsoft (among others) recently launched an Open Artificial Intelligence Energy Initiative (Pipeline Oil & Gas News, 2 February 2021). According to the press release, the initiative provides a framework for energy operators, service providers, equipment providers, and independent software vendors for energy services.

Ghoddusi *et al.* (2019) provide several examples of machine or statistical applications for energy policy. For example, machine learning techniques can be used for energy policy analysis by providing short- and long-term forecasting of energy supply and demand. Many recent papers have used statistical learning to predict smart grid electricity load, and the demand for natural gas, transport energy, and coal (Limanond *et al.* , 2011; Yang *et al.* , 2014; Bassamzadeh & Ghanem, 2017; Panapakidis & Dagoumas, 2017). These forecasting methods aid stakeholders in the development of future energy policies.

Other recent applications include using statistical learning to improve energy efficiency, energy savings, and reduce overcapacity (Azadeh *et al.* , 2007; Skiba *et al.* , 2017; Wang *et al.* , 2018; Mashhadi & Behdad, 2018; Amiri *et al.* , 2020). Similar to the current study, Mashhadi & Behdad (2018) used a shrinkage estimator to determine the effect of electronics usage on household energy demand. In terms of relevant policy implications, they point to the importance of understanding household behavior and the heterogeneous use of appliances within the home. The out-of-sample prediction features of statistical learning will become increasingly useful in the future, with greater adoption of in-home digital meters, home energy managements systems for automation, and distributed energy resources like rooftop solar PV, electric vehicles, and battery storage.

References

Alberini, A., & Filippini, M. 2011. Response of residential electricity demand to price: The effect of measurement error. *Energy Economics*, **33**(5), 889–895.

Allison, P.D. 2010. Missing Data. *Pages 631–657 of:* Marsden, P.V., & Wright, J.D. (eds), *Handbook of Survey Research*, second edn. Sommerville, MA: Emerald Group Publishing, Limited.

Amiri, S.S., Mostafavi, N., Lee, E.R., & Hoque, S. 2020. Machine learning approaches for predicting household transportation energy use. *City and Environmental Interactions*, **7**, 1–9.

Athey, S., & Imbens, G. 2017. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, **31**(2), 3–32.

Athey, S., & Imbens, G. 2019. Machine learning methods that economists should know about. *Annual Review of Economics*, **11**, 685–725.

Auffhammer, M., & Rubin, E. 2018. *Natural gas price elasticities and optimal cost recovery under consumer heterogeneity: Evidence from 300 million natural gas bills*. NBER Working Paper 24295, Accessed online September 2020, https://www.nber.org/papers/w24295?sy=295.

Ayesh, R. October 11 2019. Americans are using less electricity thanks to LED light bulbs. *Axios*.

Azadeh, A., Ghaderi, S., Anvari, M., & Saberi, M. 2007. Performance assessment of electric power generations using an adaptive neural network algorithm. *Energy Policy*, **35**(6), 3155–3166.

Baker, P., & Blundell, R. 1991. The microeconometric approach to modelling energy demand: some results for UK households. *Oxford Review of Economic Policy*, **7**(2), 54–76.

Bassamzadeh, N., & Ghanem, R. 2017. Multiscale stochastic prediction of electricity demand in smart grids using Bayesian networks. *Applied Energy*, **35**(10), 5229–5241.

Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. 2012. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, **80**(6), 2369–2429.

Belloni, A., Chernozhukov, V., & Hansen, C. 2014a. High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, **28**(2), 29–50.

Belloni, A., Chernozhukov, V., & Hansen, C. 2014b. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, **81**(2), 608–650.

Berndt, Ernst R. 1991. *The practice of econometrics: classic and contemporary*. Reading, Mass.; Don Mills, Ont.: Addison-Wesley Publishing Company.

Bernstein, M.A., & Griffin, J. 2005. *Regional differences in the price-elasticity of demand for energy*. RAND Technical Report, Accessed online September 2020, https://www.rand.org/pubs/technical_reports/TR292.html.

Bohi, Douglas R, & Zimmerman, Mary Beth. 1984. An update on econometric studies of energy demand behavior. *Annual Review of Energy*, **9**(1), 105–154.

Chernozhukov, V., Hansens, C., & Spindler, M. 2015. Post-section and post-regularization inference in linear models with many controls and instruments. *American Economic Review*, **105**(5), 486–490.

Davis, L. 2017. Evidence of a decline in electricity use by US households. *Economics Bulletin*, **37**(2), 1098–1105.

de la Rue du Can, S., Leventis, G., Phadke, A., & Gopal, A. 2014. Design of incentive programs for accelerating penetration of energy-efficient appliances. *Energy Policy*, **72**, 56–66.

Deryugina, T., MacKay, A., & Reif, J. 2017. *The long-run elasticity of demand: Evidence from municipal electric aggregation*. Accessed online September 2020, https://www.econ.pitt.edu/sites/default/files/Deryugina.Electricity%20Aggregation.pdf.

Dubin, J., & McFadden, D. 1984. An econometric analysis of residential electric appliance holdings and consumption. *Econometrica*, **52**(2), 345–362.

Economic Policy Institute. 2019. *Slowdown in household income growth continues in 2018*. Accessed pnline August 2020, https://www.epi.org/blog/slowdown-in-household-income-growth-continues-in-2018/.

Electric Power Research Institute. 2008. *Price elasticity of demand for electricity: A primer and synthesis*. Accessed online September 2020, https://www.epri.com/research/products/1016264.

Fouquet, R. 2014. Long-run demand for energy services: Income and price elasticities over two hundred years. *Review of Environmental Economics and Policy*, **8**(2), 187–207.

Gelman, A., & Imbens, G. 2013. *Why ask Why? Forward Causal Inference and Reverse Causal Questions*. Accessed online October 2020, https://www.nber.org/system/files/working_papers/w19614/w19614.pdf.

Ghoddusi, H., Creamer, G.C., & Rafizadeh, N. 2019. Machine learning in energy economics and finance: A review. *Energy Economics*, **81**, 709–727.

Hausman, C., & Kellog, R. 2015. Welfare and Distributional Implications of Shale Gas. *Brookings Papers on Economic Activity*, **Spring 2015**, 71–125.

Huebner, G., Hamilton, I., Chalabi, Z., Shipworth, D., & Oreszczyn, T. 2015. Explaining domestic energy consumption–the comparative contribution of building factors, socio-demographics, behaviours and attitudes. *Applied Energy*, **159**, 589–600.

International Energy Agency. 2019. *Multiple Benefits of Energy Efficiency*. Accessed online October 2020, https://www.iea.org/reports/multiple-benefits-of-energy-efficiency.

Ironmonger, D.S., Aitken, C.K., & Erbas, B. 1995. Economies of scale in energy use in adult-only households. *Energy Economics*, **17**(4), 301–310.

James, G., Witten, D., Hastie, T., & Tibshirani, R. 2013. *An introduction to statistical learning: With applications in R*. New York, NY: Springer Science+Business Media.

Jordan, M.I., & Mitchell, T.M. 2015. Machine learning: Trends, perspectives, and prospects. *Science*, **349**(6245), 255–260.

Limanond, T., Jomnonkwao, S., & Srikaew, A. 2011. Projection of future transport energy demand of Thailand. *Energy Policy*, **39**(5), 2754–2763.

Lovins, A.B. 1988. Energy saving resulting from the adoption of more efficient appliances. *The Energy Journal*, **9**(2), 155–162.

Maddala, G.S., Trost, R.P., Li, H., & Joutz, F. 1997. Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business & Economic Statistics*, **15**(1), 90–100.

Mashhadi, A.R., & Behdad, S. 2018. Discriminant effects of consumer electronics use-phase attributes on household energy prediction. *Energy Policy*, **118**, 346–355.

McGinty, J.C. October 11 2019. Americans are no longer gluttons for electricity – thank the LED bulb. *The Wall Street Journal*.

Nadel, S., Elliott, N., & Langer, T. 2015. *Energy efficiency in the United States: 35 years and counting*. Accessed online September 2020, https://www.aceee.org/sites/default/files/publications/researchreports/e1502.pdf.

Open Climate Fix. 2020. *Open Climate Fix: Using computers to fix climate change*. Accessed online December 2020, https://openclimatefix.org/#about.

Panapakidis, I.P., & Dagoumas, A.S. 2017. Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model. *Energy*, **118**, 231–245.

Paul, A.C., Myers, E.C., & Palmer, K.L. 2009. *A partial adjustment model of U.S. electricity demand by region, season, and sector*. RFF Discussion Paper 08-50, Accessed online September 2020, https://www.rff.org/publications/working-papers/a-partial-adjustment-model-of-us-electricity-demandby-region-season-and-sector/.

Pipeline Oil & Gas News. 2 February 2021. *Shell, C3 AI, Baker Hughes, and Microsoft launch open AI energy initiative*. Accessed online February 2020, https://www.pipelineoilandgasnews.com/techpeople/tools-and-tech/2021/february/shell-c3-ai-baker-hughes-and-microsoft-launch-open-ai-energy-initiative/.

Ros, A.J. 2017. An econometric assessment of electricity demand in the United States using utility-specific panel data and the impact of retail competition on prices. *The Energy Journal*, **38**(4), 73–99.

Ryan, D.L., & Plourde, A. 2009. Empirical modeling of energy demand. *Pages 112–143 of:* Evans, J., & Hunt, L. (eds), *International Handbook on the Economics of Energy*. Cheltenham: Edward Elgar.

Skiba, M., Mrówczy'nska, M., & Bazan-Krzywosza'nska, A. 2017. Modeling the economic dependence between town development policy and increasing energy effectiveness with neural networks. Case study: the town of Zielona Góra. *Applied Energy*, **188**, 356–366.

Solon, G., Haider, S.J., & Wooldridge, J.M. 2015. What are we weighting for? *The Journal of Human Resources*, **50**(2), 301–316.

StataCorp. 2019. *Stata lasso reference manual: Release 16*. Accessed online November 2020, https://www.stata.com/manuals/lasso.pdf.

Stewart, M. 24 March 2019. *The actual difference between statistics and machine learning*. Accessed online December 2020, https://towardsdatascience.com/the-actual-difference-between-statistics-and-machine-learning-64b49f07ea3.

Taylor, L.D. 1975. The demand for electricity: A survey. *The Bell Journal of Economics*, **6**(1), 74–110.

U.S. Bureau of Labor Statistics. 2020a. *Consumer Expenditure Survey*. Table 1502. Accessed online July 2020, https://www.bls.gov/cex/2018/combined/cucomp.pdf.

U.S. Bureau of Labor Statistics. 2020b. *CPI for All Urban Consumers: Energy in U.S. city average, all urban consumers, not seasonally adjusted*. Accessed online August 2020, https://data.bls.gov/pdq/SurveyOutputServlet.

U.S. Census Bureau. 2020a. *Characteristics of New Housing*. Accessed online September 2020, https://www.census.gov/construction/chars/.

U.S. Census Bureau. 2020b. *State Population Totals and Components of Change: 2010-2019*. Accessed online August 2020, https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html.

U.S. Energy Information Administration. 2018. *Electricity: Detailed State Data*. Accessed online August 2020, https://www.eia.gov/electricity/data/state/.

U.S. Energy Information Administration. 2019a. *Detailed State Data: Average Price by State by Provider (EIA-861)*. Accessed online August 2020, https://www.eia.gov/electricity/data/state/.

U.S. Energy Information Administration. 2019b. *Electric Sales, Revenue, and Average Price*. Accessed online August 2020, https://www.eia.gov/electricity/sales_revenue_price/.

U.S. Energy Information Administration. 2019c. *Use of energy explained: Energy use in homes*. Accessed online June 2020, https://www.eia.gov/energyexplained/use-of-energy/homes.php.

U.S. Energy Information Administration. 2020a. *Monthly Energy Review*. Table 2.1. Accessed online July 2020, https://www.eia.gov/totalenergy/data/monthly/.

U.S. Energy Information Administration. 2020b. *Natural Gas: Natural Gas Prices*. Accessed online August 2020, https://www.eia.gov/dnav/ng/ng_pri_sum_a_EPG0_PRS_DMcf_a.htm.

U.S. Energy Information Administration. 2020c. *Natural Gas: Natural Gas Summary*. Accessed online August 2020, https://www.eia.gov/dnav/ng/ng_sum_lsum_a_EPG0_vrs_mmcf_a.htm.

U.S. Energy Information Administration. 2020d. *Use of energy explained: Energy use in homes*. Access online August 2020, https://www.eia.gov/energyexplained/use-of-energy/homes.php.

U.S. Energy Information Administration. 2021a. *Petroleum & other liquids: U.S. No. 2 heating oil residential price*. Accessed online January 2021, https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=M_EPD2F_PRS_NUS_DPG&f=M.

U.S. Energy Information Administration. 2021b. *Petroleum & other liquids: U.S. residential propane prices*. Accessed online January 2021, https://www.eia.gov/dnav/pet/pet_pri_wfr_a_EPLLPA_PRS_dpgal_m.htm.

Wang, D., Wang, Y., Song, X., & Liu, Y. 2018. Coal overcapacity in China: Multiscale analysis and prediction. *Energy Economics*, **70**, 244–257.

Witherspoon, S., & Fadrhonc, W. 26 February 2019. *Machine learning can boost the value of wind energy*.

Yang, J.H., Wu, C.Y., & Zhang, G.M. 2014. Coal demand prediction in Shandong Province based on artificial firefly wavelet neural network. *Advanced Material Research*, **962**, 1931–1935.

Young, D. 2008. When do energy-efficient appliances generate energy savings? Some evidence from Canada. *Energy Policy*, **36**(1), 34–46.

Acknowledgements

The findings and conclusions in this presentation are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.

This research was supported by the intramural research program of the USDA's Economic Research Service and Carnegie Mellon University.

Table 1: Mean and standard deviation for selected common variables across the surveys in 2001, 2005, 2009, and 2015

	2001		2005		2009		2015	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Type of home (categorical indicator)								
Mobile home	0.08	0.26	0.07	0.25	0.05	0.22	0.05	0.22
Detached home	0.63	0.48	0.66	0.47	0.67	0.47	0.68	0.47
Attached home	0.09	0.29	0.08	0.27	0.07	0.26	0.09	0.29
Apt Bldg 2-4 units	0.09	0.28	0.06	0.24	0.07	0.25	0.05	0.23
Apt Bldg 5 or more units	0.12	0.32	0.14	0.34	0.14	0.35	0.12	0.33
Total square feet (thousands)	2.11	1.26	2.30	1.51	2.16	1.27	2.04	1.09
Total BTUs (thousands)	97.43	48.95	102.02	52.30	89.61	46.35	76.80	39.97
Number of beds	2.76	0.96	2.79	1.01	2.90	1.01	2.88	1.02
Number of baths	1.50	0.63	1.57	0.68	1.69	0.71	1.75	0.70
Own home (binary)	0.28	0.45	0.29	0.46	0.30	0.46	0.28	0.45
Home temperature	72.07	7.81	72.26	7.53	69.84	4.25	67.11	14.95
Air conditioning (binary)	0.79	0.41	0.84	0.36	0.85	0.36	0.88	0.32
Head of household age	49.89	18.28	49.16	17.04	49.49	16.49	52.65	16.99
Head of household employed (binary)	0.53	0.50	0.53	0.50	0.51	0.50	0.48	0.50
Numbers of household members	2.63	1.43	2.67	1.48	2.70	1.50	2.59	1.41
Energy expenditures (US$ thousands)								
Electricity	0.96	5.46	1.12	0.67	1.39	0.80	1.43	0.74
Natural gas	0.42	0.47	0.78	0.48	0.46	0.53	0.33	0.38
Heating oil	0.09	0.27	1.55	0.60	0.09	0.38	0.06	0.31
Propane	0.07	0.24	0.87	0.56	0.08	0.343	0.06	0.26
Kerosene	0.007	0.06	0.14	0.172	–	–	–	–
Number of obs.	3,609		3,501		8,172		4,389	

Table 2: Least squares estimates for household energy demand

(Dependent variable: Total household energy consumption (BTUs))

Variables	**2001**	**2005**	**2009**	**2015**
ENERGY PRICE ELASTICITIES				
Electricity	-0.0793*	-0.2022***	-0.1239***	-0.2587***
	(0.0419)	(0.0449)	(0.0304)	(0.0389)
Natural gas	-0.3461***	-0.1793***	-0.4392***	-0.3010***
	(0.0344)	(0.0375)	(0.0274)	(0.0269)
Heat oil	-0.1877	0.2231	-0.9177**	0.5262
	(0.3021)	(0.6799)	(0.4023)	(0.9773)
Propane	-0.5609***	-0.2937***	-0.3305***	-0.3748***
	(0.0782)	(0.0761)	(0.0957)	(0.1243)
Kerosene	-0.9790	–	–	–
	(1.7718)			
HOUSEHOLD CHARACTERISTICS				
Total square feet	0.1339***	0.0970***	0.1514***	0.1403***
	(0.0164)	(0.0166)	(0.0119)	(0.0154)
No. household members	0.0697***	0.0734***	0.0638***	0.0754***
	(0.0051)	(0.0058)	(0.0033)	(0.0043)
Householder age	0.0003	0.0012**	-0.0002	0.0006
	(0.0004)	(0.0005)	(0.0003)	(0.0004)
Constant	12.22***	11.37***	15.57***	11.16***
	(0.8213)	(0.5888)	(0.5410)	(0.9333)
Number of obs.	3,077	3,002	7,015	3,823
R-squared	0.679	0.640	0.682	0.684

Notes: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Scale elasticity estimates based on total annual electricity expenditures ($US)

Dependent variable: Annual expenditures on electricity

Panel A: Regression estimates				
Members	**(2001)**	**(2005)**	**(2009)**	**(2015)**
2	202.36***	263.20***	233.04***	296.56***
	(20.54)	(25.25)	(20.13)	(25.38)
3	315.45***	429.98***	415.48***	395.67***
	(25.67)	(30.87)	(25.45)	(32.61)
4	430.12***	494.17***	546.93***	521.81***
	(29.86)	(37.40)	(29.13)	(38.56)
5	539.03***	646.04***	593.44***	718.99***
	(47.66)	(54.24)	(42.20)	(57.65)
6	535.96***	531.61***	689.51***	654.14***
	(73.49)	(73.74)	(61.78)	(80.63)
7	477.96***	666.17***	937.76***	1109.68***
	(180.35)	(140.94)	(127.61)	(185.80)
8	534.04**	609.61***	888.54***	755.05***
	(232.39)	(124.79)	(169.27)	(239.55)
9	537.13**	1,007.83***	716.62***	-385.32***
	(230.68)	(331.99)	(219.09)	(21.32)
10	1,855.21	412.97	1,019.51	–
	(1,161.18)	(307.67)	(652.80)	

Panel B: Contrasted predicted values				
Members	**(2001)**	**(2005)**	**(2009)**	**(2015)**
(1 vs 2)	-202.36***	-263.20***	-233.04***	-296.56***
	(20.54)	(25.25)	(20.13)	(25.38)
(2 vs 3)	-113.09***	-166.78***	-182.44***	-99.11***
	(25.29)	(31.39)	(25.38)	(31.04)
(3 vs 4)	-114.67***	-64.19	-131.45***	-126.14***
	(32.78)	(41.39)	(32.72)	(42.38)
(4 vs 5)	-108.92**	-151.87**	-46.52	-197.18***
	(51.98)	(60.84)	(46.40)	(63.64)
(5 vs 6)	3.08	114.43	-96.07	64.85
	(85.05)	(88.07)	(71.69)	(95.10)
(6 vs 7)	57.99	-134.56	-248.25	-455.54**
	(193.34)	(156.85)	(140.07)	(200.83)
(7 vs 8)	-56.08	56.56	49.22	354.63
	(293.17)	(186.42)	(211.14)	(301.92)
(8 vs 9)	-3.09	-398.22	171.93	1,140.37***
	(326.58)	(353.79)	(276.25)	(238.68)
(9 vs 10)	-1,318.08	594.86	-302.89	–
	(1,183.53)	(451.71)	(688.07)	

Notes: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Scale elasticity estimates based on total annual natural gas expenditures ($US)

Dependent variable: Annual expenditures on natural gas

Panel A: Regression estimates

Members	**(2001)**	**(2005)**	**(2009)**	**(2015)**
2	44.10**	115.76***	86.31***	15.81
	(19.87)	(23.99)	14.26	(13.96)
3	74.66***	141.21***	126.26**	63.65***
	(25.00)	(31.56)	18.04	(18.43)
4	73.64***	109.91***	131.71***	42.50**
	(25.70)	(31.22)	19.32	(18.95)
5	99.19***	217.10***	124.67***	27.81
	(37.73)	(46.90)	25.19	(25.13)
6	136.59**	162.68***	173.46***	52.02
	(63.18)	(59.97)	39.93	(46.23)
7	324.00**	257.28*	132.46**	20.21
	(131.10)	(132.15)	60.61	(83.26)
8	275.26	357.52**	187.65*	141.31
	(189.67)	(140.64)	97.64	(100.81)
9	-64.53	474.68	191.96	182.20***
	(158.69)	(355.88)	144.45	(11.64)
10	-265.09	-84.28*	163.86***	–
	(269.00)	(49.16)	63.05	

Panel B: Contrasted predicted values

Members	**(2001)**	**(2005)**	**(2009)**	**(2015)**
(1 vs 2)	-44.10**	-115.76***	-86.31***	-15.81
	(19.87)	(23.99)	(14.26)	(13.96)
(2 vs 3)	-30.56	-25.45	-39.95**	-47.85***
	(24.33)	(30.90)	(17.63)	(17.93)
(3 vs 4)	1.02	31.31	-5.46	21.15
	(29.06)	(36.63)	(21.83)	(21.96)
(4 vs 5)	-25.55	-107.19**	7.04	14.69
	(40.13)	(50.53)	(27.92)	(27.73)
(5 vs 6)	-37.40	54.42	-48.79	-24.21
	(70.25)	(71.94)	(44.74)	(50.50)
(6 vs 7)	-187.42	-94.60	41.00	31.80
	(143.86)	(142.79)	(71.08)	(94.07)
(7 vs 8)	48.75	-100.25	-55.19	-121.20
	(229.56)	(191.28)	(114.07)	(129.84)
(8 vs 9)	339.78	-117.16	-4.32	-40.89
	(246.29)	(382.10)	(173.82)	(100.23)
(9 vs 10)	200.57	558.95	28.20	–
	(310.98)	(359.27)	(156.74)	

Notes: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

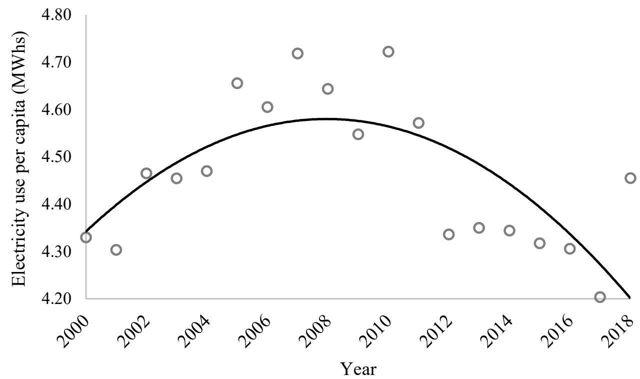
Table 5: Double-selection linear model estimates of price elasticities of demand

	2001					2005			
	(1)	**(2)**	**(3)**	**(4)**		**(1)**	**(2)**	**(3)**	**(4)**
Variables	**OLS**	**Plugin**	**CV**	**Adapt**	**Variables**	**OLS**	**Plugin**	**CV**	**Adapt**
Electricity	-0.0793*	-0.1452***	-0.1568***	-0.1541***	Electricity	-0.2022***	-0.1279**	-0.1678***	-0.1641***
	(0.0419)	(0.0486)	0.0478	(0.0476)		(0.0449)	(0.0508)	(0.0488)	(0.0481)
Natural gas	-0.3461***	-0.3451***	-0.3161***	-0.3118***	Natural gas	-0.1793***	-0.2157***	-0.2309***	-0.2282***
	(0.0344)	(0.0368)	(0.0374)	(0.0372)		(0.0375)	(0.0403)	(0.0398)	(0.0393)
Heating oil	-0.1877	0.2852	0.3032	0.2908	Heating oil	0.2231	-0.0965	-0.1398	-0.0440
	(0.3021)	(0.3181)	(0.3160)	(0.3100)		(0.6799)	(0.7589)	(0.7136)	(0.7116)
Propane	-0.5609***	-0.1125	-0.2974***	-0.2721***	Propane	-0.2937***	0.0254	-0.1476	-0.1372
	(0.0782)	(1.2650)	(0.1003)	(0.0978)		(0.0761)	(0.0916)	(0.0980)	(0.0980)
MSE	0.1075	0.1517	0.0800	0.0995	MSE	0.1092	0.1310	0.0926	0.0951
R-squared	0.6257	0.4742	0.6438	0.6533	R-squared	0.5970	0.5111	0.6500	0.6453
OS obs.	769	765	594	760	OS obs.	750	746	731	746

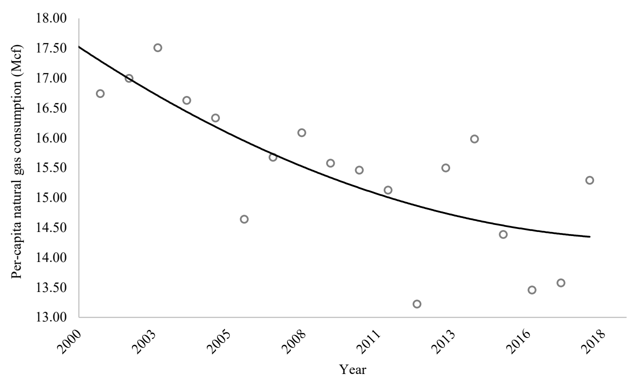
	2009					2015			
	(1)	**(2)**	**(3)**	**(4)**		**(1)**	**(2)**	**(3)**	**(4)**
Variables	**OLS**	**Plugin**	**CV**	**Adapt**	**Variables**	**OLS**	**Plugin**	**CV**	**Adapt**
Electricity	-0.1239***	-0.1559***	-0.1693***	-0.1683***	Electricity	-0.2587***	-0.2992***	-0.2835***	-0.2860***
	(0.0304)	(0.0294)	(0.0288)	(0.0288)		(0.0389)	(0.0397)	(0.0403)	(0.0402)
Natural gas	-0.4392***	-0.4256***	-0.4145***	-0.4142***	Natural gas	-0.3010***	-0.3089***	-0.3148***	-0.3150***
	(0.0274)	(0.0270)	(0.268)	(0.0268)		(0.0269)	(0.0281)	(0.0282)	(0.0282)
Heating oil	-0.9177**	-0.6480*	-0.7490**	-0.7599**	Heating oil	0.5262	-0.2746	-0.5121	-0.4466
	(0.4023)	(0.3675)	(0.3735)	(0.3730)		(0.9773)	(0.9188)	(0.9582)	(0.9701)
Propane	-0.3305***	-0.2928***	-0.2150	-0.2132**	Propane	-0.3748***	-0.2969**	-0.2521**	-0.2672**
	(0.0957)	(0.0995)	(0.0958)	(0.0958)		(0.1243)	(0.1212)	(0.1162)	(0.1190)
MSE	0.1067	0.1031	0.0849	0.0855	MSE	0.1104	0.0971	0.0858	0.0863
R-squared	0.6199	0.6290	0.6944	0.6923	R-squared	0.5715	0.6556	0.6956	0.6940
OS obs.	1,754	1,747	1,747	1,747	OS obs.	956	866	866	866

Notes: "MSE" denotes mean-squared-error; "OS obs" denotes out-of-sample observation; "Plugin" denotes a plug-in value method to estimate the lasso tuning parameter; "CV" denotes cross-validation method to estimate the lasso tuning parameter; "Adapt" denotes an adaptive method to estimate the lasso tuning parameter; and "OLS" denotes ordinary least squares. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 1: U.S. average, per-capita residential electricity and natural gas consumption (2000-2018)



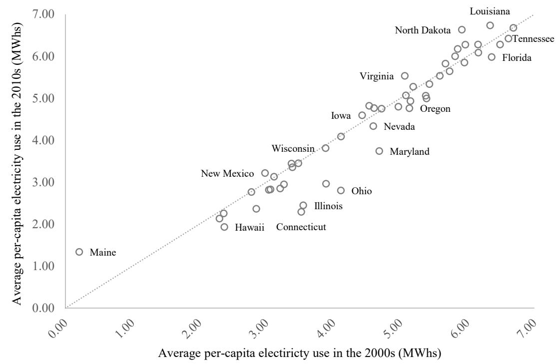
(a) Electricity



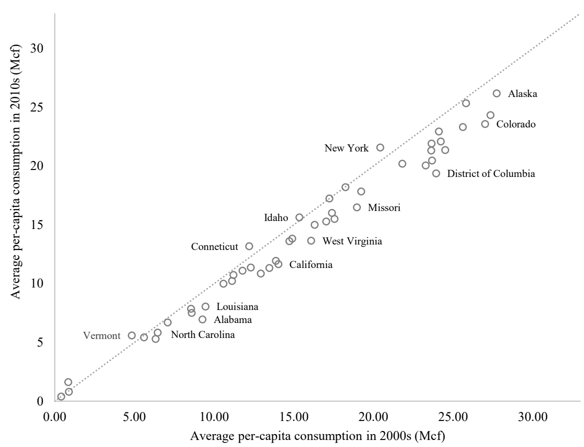
(b) Natural gas

Source: Constructed by the authors using residential electricity (U.S. Energy Information Administration, 2019b) and natural gas (U.S. Energy Information Administration, 2020c) consumption data and population statistics (U.S. Census Bureau, 2020b).

Figure 2: State-level average per-capital residential electricity and natural gas consumption by decade (2000-2018)



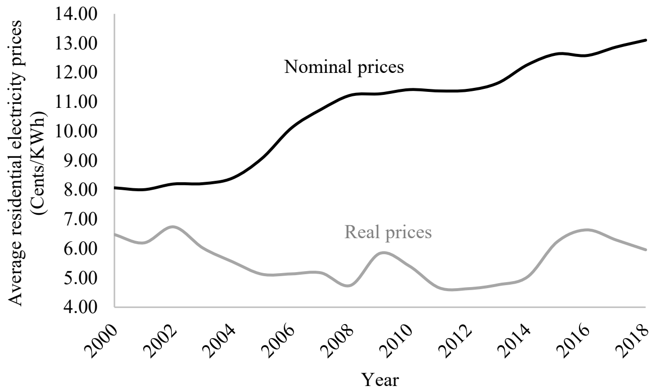
(a) Electricity



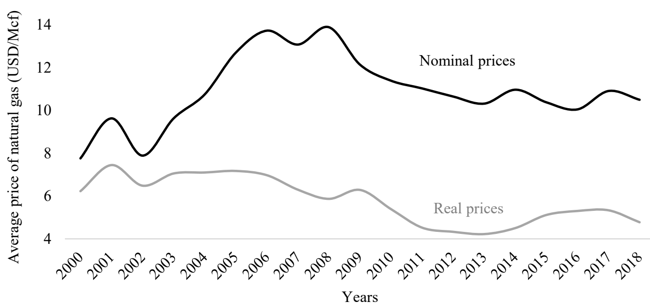
(b) Natural gas

Source: Constructed by the authors using residential electricity (U.S. Energy Information Administration, 2019b) and natural gas (U.S. Energy Information Administration, 2020c) consumption data and population statistics (U.S. Census Bureau, 2020b).

Figure 3: U.S. average residential electricity and natural gas prices (2000-2018)



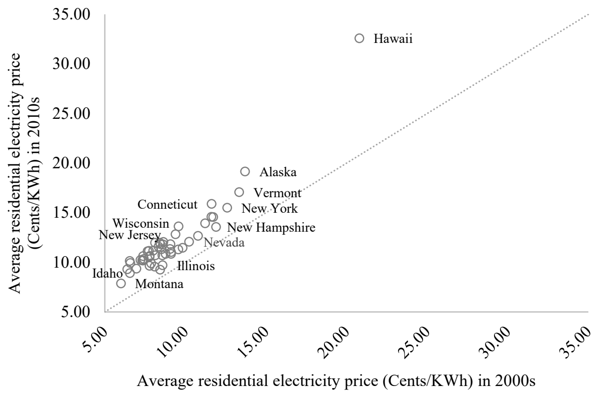
(a) Electricity



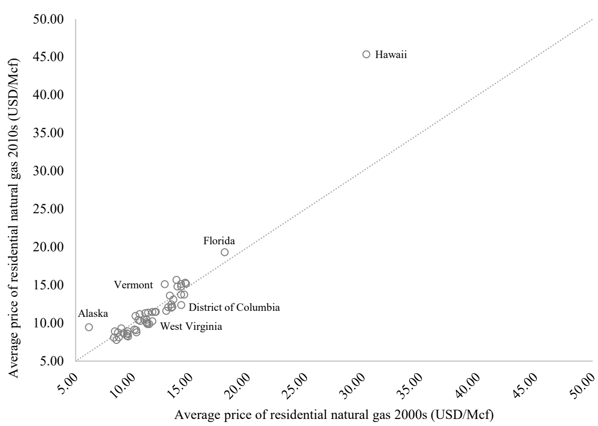
(b) Natural gas

Source: Constructed by the authors using residential electricity (U.S. Energy Information Administration, 2019a) and natural gas (U.S. Energy Information Administration, 2020b) price data, and the adjusted for inflation using the Consumer Price Index (U.S. Bureau of Labor Statistics, 2020b).

Figure 4: State-level average per-capital residential electricity and natural gas prices by decade (2000-2018)



(a) Electricity



(b) Natural Gas

Source: Constructed by the authors using residential electricity (U.S. Energy Information Administration, 2019a) and natural gas (U.S. Energy Information Administration, 2019a) price data.

A Variables across RECS survey years

This list *only* contains the common variables collected across the surveys examined within this study. The actual number of variables collected varies by each of the surveys, and the total variables are substantially larger than the list provided in Table 6,

Table 6: List and description of data collected by the RECS survey

Beginning of Table

Number	Variable	Data Type	Description
1	doeid	Count	Unique id created by US DOE
2	year	Count	Year of RECS survey
3	nweight	Count	Survey weight created by US DOE
4	totusqft	Count	Total square feet of heated space
5	attic	Indicator	Housing unit has attic space
6	atticfin	Indicator	At least part of the attic is completed
7	attcheat	Indicator	Attic space is heated
8	protherm	Indicator	Housing unit has programmable thermostat
9	thermain	Indicator	Housing unit has thermostat to control heating
10	numberac	Count	Number of window A/C units in housing unit
11	usecenac	Categorical	How housing unit uses A/C during the summer
12	numfrig	Count	Number of separate refrigerators in the house
13	typerfr1	Categorical	Type of first refrigerator
14	sizrfri1	Count	Size of first refrigerator
15	elfood	Indicator	Housing unit uses electricity for food prep
16	dishwash	Indicator	Housing unit has dishwasher
17	dwashuse	Indicator	Housing unit uses dishwasher
18	cwasher	Indicator	Housing unit uses clothes washing machine
19	fuelpool	Categorical	Type of fuel used to heat pool
20	fueltub	Categorical	Fuel used to heat hot tub
21	recbath	Indicator	Housing unit has hot tub

Continued on next page

Continuation of Table 6

Number	Variable	Data Type	Description
22	equipaux	Indicator	Use of any other type of equipment to heat home
23	btuelrfg	Count	BTUs for refrigeration
24	dollpwth	Count	Estimated dollars spent on propane for water heating
25	ice	Indicator	Refrigerator contains through-the-door ice maker
26	regionc	Categorical	Census region
27	prkgplc1	Indicator	Housing unit has parking garage or carport
28	topfront	Indicator	Washing machine loading direction
29	tvcolor	Indicator	Housing unit has color television
30	temphome	Count	Self-reported value of thermostat set during the day
31	wwacage	Count	Age of the oldest window- or wall-unit A/C
32	elwater	Indicator	Electricity is used for water heating
33	wdwater	Indicator	Wood is used for water heating
34	periodng	Categorical	Period of time natural gas is available
35	btungsph	Count	BTUs of natural gas for heating
36	dollarlp	Count	Estimated dollars spent on propane per year
37	typehuq	Categorical	Type of housing unit
38	kownrent	Categorical	Housing unit owned or rented (1 = own; 2 = rent)
39	typerfr2	Categorical	Type of second refrigerator
40	washtemp	Count	Temperature setting for wash cycle
41	notmoist	Indicator	Electric dehumidifier
42	tempnite	Count	Estimated thermostat setting during the night hours
43	door1sum	Count	Number of sliding glass doors in the housing unit
44	ugwarm	Indicator	Natural gas used to heat interior of housing unit
45	solother	Indicator	Solar energy used for other purposes
46	periodlp	Categorical	Period of time propane is available
47	btulpsph	Count	BTUs of propane used for heating
48	dolngsph	Count	Dollars on natural gas for space heating
49	cellar	Indicator	Housing unit contains finished cellar or basement
50	stove	Indicator	Housing unit has separate built-in range or stove
51	stories	Count	Number of stories within housing unit
52	stoven	Indicator	Housing unit has burners and at least one oven

Continued on next page

Continuation of Table 6			
Number	Variable	Data Type	Description
53	sizrfri2	Count	Size of the second refrigerator
54	rnsetemp	Count	Water temperature setting for rinse cycle in washer
55	moisture	Indicator	Electric humidifier is used during the winter
56	fuelh2o	Categorical	Type of fuel used for water heating
57	windows	Count	Number of windows within housing unit
58	ugwater	Indicator	Natural gas used for water heating
59	fopay	Categorical	Who pays for fuel oil?
60	totalbtu	Count	BTUs of energy consumed
61	btuelwth	Count	BTUs of electricity for water heat
62	dolfosph	Count	Estimated dollars spent on fuel oil for space heating
63	ncombath	Count	Number of completed bathrooms in housing unit
64	ovenuse	Indicator	Housing unit uses oven for food preparation
65	uprtfrzr	Categorical	Model of freezer
66	dryruse	Count	How often clothing dryer is used
67	dntheat	Indicator	Heating equipment used during the year
68	aircond	Indicator	Housing unit contains central air conditioning
69	useel	Indicator	Housing unit uses electricity
70	lpwarm	Indicator	Housing unit uses propane for heating
71	athome	Indicator	At home during the week
72	cufeetng	Count	Cubic feet of natural gas consumption
73	btulpwth	Count	BTUs of propane used for water heating
74	dolngwth	Count	Dollars spent on propane for water heating
75	dryer	Indicator	Clothes dryer is used within housing unit
76	lpother	Indicator	Propane is used for other purposes in housing unit
77	employhh	Categorical	Head of household is employed
78	btufo	Count	BTUs of fuel oil consumed per year
79	usesolar	Indicator	Housing unit has solar energy
80	fowater	Indicator	Housing unit uses fuel oil for water heating
81	moneypy	Categorical	Income of head of the household
82	gallonlp	Count	Gallons of propane consumed per year
83	dollarel	Count	Total dollars spent on electricity by year

Continued on next page

Continuation of Table 6

Number	Variable	Data Type	Description
84	dolelcol	Count	Estimated dollars spent on electricity for A/C
85	dollarng	Count	Total dollars spent on natural gas per year
86	dolelrfg	Count	Estimated dollars spent on refrigeration per year
87	dollarfo	Count	Estimated dollars spent on fuel oil per year
88	elwarm	Indicator	Housing unit uses electricity for water heating
89	foother	Indicator	Fuel oil is used for other purposes
90	btulp	Count	Total BTUs of propane consumed
91	wdwarm	Indicator	Housing unit uses wood for internal heating
92	periodel	Categorical	Period of time electricity is available
93	btuelsph	Indicator	Estimated BTUs of electric for space heating
94	bedrooms	Count	Number of bedrooms within housing unit
95	amtmicro	Count	How often the microwave is used on a weekly basis
96	agefrzr	Count	Age of the main freezer within the housing unit
97	swimpool	Indicator	Housing unit has a swimming pool
98	equipage	Count	Age of equipment (other than central heat) used for heating
99	cenachp	Indicator	Central A/C system is a heat pump
100	uselp	Indicator	Housing unit uses propane fuel
101	lpcook	Indicator	Housing unit uses propane for food preparation
102	hhage	Count	Age of housing unit
103	gallonfo	Count	Estimated gallons of fuel oil consumed per year
104	division	Categorical	Census division
105	gargheat	Indicator	Parking garage is heated
106	agerfri1	Count	Age of the first refrigerator
107	washload	Count	Number of loads of laundry washed per week
108	swampcol	Indicator	Housing unit contains evaporative or swamp cooler
109	tempgone	Count	Self-reported thermostat setting when away from the home
110	usewwac	Categorical	How housing unit uses window-unit A/C during the summer
111	elother	Indicator	Electricity used for other purposes
112	solwater	Indicator	Solar electricity used for water heating
113	periodfo	Categorical	Period of time fuel oil is available
114	btufosph	Count	Estimated BTUs of fuel oil for space heating

Continued on next page

Continuation of Table 6

Number	Variable	Data Type	Description
115	dolelsph	Count	Estimated dollars spent on electricity for space heating
116	baseheat	Indicator	Basement or crawl space interior is heated
117	oven	Indicator	Housing unit has separate built-in oven
118	numfreez	Count	Number of separate freezer unit within housing unit
119	dryrfuel	Categorical	Type of fuel used for clothing dryer
120	fuelheat	Categorical	Main type of fuel used for central heating of housing unit
121	wheatage	Count	Age of the main water heater within the housing unit
122	drafty	Count	How often is the housing unit drafty during the winter
123	ugoth	Indicator	Natural gas is used for other purposes
124	nhsldmem	Count	Number of persons residing within the housing unit
125	btuel	Count	Estimated BTUs of electricity consumed per year
126	btufowth	Count	Estimated BTUs of fuel oil used for water heating
127	dolelwth	Count	Estimated dollars spent on electricity for water heating
128	agerfri2	Count	Age of the second refrigerator within the housing unit
129	internet	Indicator	Housing unit has access to internet
130	wheatsiz	Count	Size of main water heater within housing unit
131	adqinsul	Categorical	Housing unit has adequate insulation
132	ugcook	Indicator	Natural gas is used for food preparation
133	lpgpay	Categorical	Who pays for propane?
134	kwh	Count	Estimated KWhs of electricity consumption per year
135	btungwth	Count	Estimated BTUs of natural gas used for water heating
136	dollpsph	Count	Estimated dollars spent on propane for space heating
137	nhafbath	Count	Number of half bathrooms within housing unit
138	micro	Indicator	Housing unit has microwave
139	sizfreez	Count	Size of main freezer within housing unit
140	numcfan	Count	Number of ceiling fans within housing unit
141	equipm	Categorical	Type of heating equipment used for heating housing unit
142	cooltype	Categorical	Type of cooling equipment used for cooling housing unit
143	useng	Indicator	Housing unit uses natural gas
144	lpwater	Indicator	Housing unit uses propane for water heating
145	hhsex	Binary	Head of household identifies as male or female

Continued on next page

Continuation of Table 6			
Number	Variable	Data Type	Description
146	btung	Count	Estimated BTUs of natural gas consumed per year
147	btuelcol	Count	Estimated BTUs of electricity for cooling house
148	dolfowth	Count	Estimated dollars spent on fuel oil for water heating
149	othrooms	Count	Number of other rooms (not bedrooms) in housing units
150	nummeal	Count	How often meals prepared within the home per week
151	pool	Indicator	Swimming pool is heated
152	agecenac	Count	Age of central A/C system in housing unit
153	usefo	Indicator	Housing unit uses fuel oil

B Full set of linear regression estimates

Table 7: Full estimates for the (first generation) linear regression models

(Dependent variable: Total household-level energy consumption (BTUs))				
Variables	**(2001)**	**(2005)**	**(2009)**	**(2015)**
ENERGY PRICE ELASTICITIES				
Electricity	-0.0793*	-0.2022***	-0.1242***	-0.2587***
	(0.0419)	(0.0449)	(0.0304)	(0.0389)
Natural gas	-0.3461***	-0.1793***	-0.4400***	-0.3010***
	(0.0344)	(0.0375)	(0.0274)	(0.0269)
Heating oil	-0.1877	0.2231	-0.9456**	0.5262
	(0.3021)	(0.6799)	(0.4006)	(0.9773)
Propane	-0.5609***	-0.2937***	-0.3343***	-0.3748***
	(0.0782)	(0.0761)	(0.0954)	(0.1243)
Kerosene	-0.97902			
	(1.7718)			

Table 7 – continued from previous page

Variables	(2001)	(2005)	(2009)	(2015)
HOUSEHOLD CHARACTERISTICS				
No. hshld members	0.0697***	0.0734***	0.0637***	0.0754***
	(0.0051)	(0.0058)	(0.0032)	(0.0043)
Householder age	0.0003	0.0011**	-0.0002	0.0006
	(0.0004)	(0.0005)	(0.0003)	(0.0004)
Total square feet	0.1339***	0.0970***	0.1514***	0.1403***
	(0.0164)	(0.0165)	(0.0120)	(0.0154)
HDD	4.4 x 10-5***	0.0001***	0.0001***	0.0001***
	(9.8 x 10-6)	(0.0000)	(0.0000)	(0.0000)
CDD	4.2 x 10-5**	0.0001***	0.0001***	0.0001***
	(1.8 x 10-5)	(0.0000)	(0.0000)	(0.0000)
Bedrooms	0.0650***	0.0383***	0.0351***	0.0322***
	(0.0096)	(0.0096)	(0.0057)	(0.0076)
Bathrooms	0.0344***	0.0412***	0.0580***	0.0508***
	(0.0126)	(0.0137)	(0.0080)	(0.0108)
Half baths	0.01529	0.0370***	0.0352***	0.0464***
	(0.0118)	(0.0134)	(0.0079)	(0.0109)
No. refrigerators	0.0492***	0.0787***	0.0556***	0.0632***
	(0.0145)	(0.0138)	(0.0076)	(0.0086)
Other rooms	0.0495***	0.0232***	0.0216***	0.0099**
	(0.0060)	(0.0062)	(0.0035)	(0.0044)
Thermostat temp	-0.0002	0.0010	0.0025*	-0.0001
	(0.0014)	(0.0015)	(0.0014)	(0.0016)
Thermostat (away)	0.0004	-0.0003	0.0030***	0.0034***
	(0.0008)	(0.0010)	(0.0009)	(0.0013)
Thermostat (night)	0.0003	-0.0003	0.0031***	-0.0006
	(0.0011)	(0.0012)	(0.0011)	(0.0015)
DIVISION				

Table 7 – continued from previous page

Variables	(2001)	(2005)	(2009)	(2015)
2	0.0306	-0.0206	-0.0416	-0.0715
	(0.0285)	(0.0319)	(0.0254)	(0.0455)
3	-0.1009***	-0.1339***	-0.1993***	-0.1470***
	(0.0327)	(0.0344)	(0.0270)	(0.0446)
4	-0.0781**	-0.2257***	-0.2742***	-0.2692***
	(0.0348)	(0.0388)	(0.0327)	(0.0486)
5	-0.0623*	-0.1670***	-0.2316***	-0.1369***
	(0.0364)	(0.0399)	(0.0317)	(0.0493)
6	-0.0410	-0.1160**	-0.2154***	-0.0915*
	(0.0419)	(0.0469)	(0.0350)	(0.0524)
7	-0.0532	-0.1774***	-0.2060***	-0.1624***
	(0.0404)	(0.0415)	(0.0351)	(0.0531)
8	-0.1679***	-0.2373***	-0.2986***	-0.3286***
	(0.0405)	(0.0397)	(0.0491)	(0.0495)
9	-0.1928***	-0.2523***	-0.1905***	-0.2608***
	(0.0468)	(0.0506)	(0.0490)	(0.0522)
10			-0.2574***	-0.3299***
			(0.0451)	(0.0542)
TYPE OF HOME				
2	-0.0836***	-0.0632**	-0.0918***	-0.0572**
	(0.0249)	(0.0280)	(0.0191)	(0.0260)
3	-0.1471***	-0.1399***	-0.2193***	-0.2094***
	(0.0299)	(0.0339)	(0.0226)	(0.0298)
4	-0.1559***	-0.1437***	-0.1899***	-0.3023***
	(0.0329)	(0.0386)	(0.0248)	(0.0352)
5	-0.4035***	-0.3217***	-0.3421***	-0.4357***
	(0.0327)	(0.0339)	(0.0216)	(0.0308)
URBAN/RURAL				

Table 7 – continued from previous page

Variables	(2001)	(2005)	(2009)	(2015)
1			-0.0086	0.0597***
			(0.0117)	(0.0162)
2	0.0233	0.0066		
	(0.0155)	(0.0175)		
3	0.0098	0.0256		
	(0.0159)	(0.0174)		
4	0.0233	0.0859***		0.0788***
	(0.0192)	(0.0201)		(0.0194)
METROPOLITAN/MICROPOLITAN				
Micro			0.0272**	
			(0.0128)	
None			0.0310*	
			(0.0186)	
OWN/RENT				
2	-0.0008	0.0515***	0.0225*	0.0236
	(0.0184)	(0.0194)	(0.0119)	(0.0167)
3	-0.0415	0.0127	0.0097	0.0678
	(0.0550)	(0.0627)	(0.0387)	(0.0517)
YEAR HOUSE MADE				
2	-0.0629**	-0.0573*	-0.0281*	-0.0568***
	(0.0258)	(0.0304)	(0.0152)	(0.0213)
3	-0.0919***	-0.0922***	-0.0712***	-0.0788***
	(0.0200)	(0.0249)	(0.0155)	(0.0213)
4	-0.1269***	-0.1204***	-0.0729***	-0.0852***
	(0.0213)	(0.0261)	(0.0147)	(0.0193)
5	-0.1173***	-0.0937***	-0.1121***	-0.1129***
	(0.0207)	(0.0240)	(0.0156)	(0.0199)
6	-0.1174***	-0.1632	-0.0918***	-0.1071***
	(0.0338)	(0.0291)	(0.0158)	(0.0210)

Table 7 – continued from previous page

Variables	(2001)	(2005)	(2009)	(2015)
7	-0.1546***	-0.18370***	-0.1565***	-0.1300***
	(0.0223)	(0.0284)	(0.0173)	(0.0207)
8	-0.2465***	-0.1481***	-0.1514***	-0.1448***
	(0.0495)	(0.0300)	(0.0197)	(0.0290)
9	-0.2349***	-0.1811***		
	(0.0265)	(0.0279)		
10	-0.2206***	-0.1559***		
	(0.0395)	(0.0303)		
11	-0.1858***	-0.2098***		
	(0.0463)	(0.0475)		
12	-0.2763***	-0.2134***		
	(0.0685)	(0.0506)		
13	-0.5707***	-0.1712***		
	(0.1193)	(0.0656)		
INCOME				
2	-0.0301	0.1572	0.0468	0.0386**
	(0.0414)	(0.1317)	(0.0476)	(0.0177)
3	-0.0171	0.0737	0.0285	0.0652***
	(0.0474)	(0.1163)	(0.0424)	(0.0186)
4	0.0188	0.0964	0.0061	0.0784***
	(0.0487)	(0.1161)	(0.0361)	(0.0199)
5	0.0278	0.1364	0.0085	0.0971***
	(0.0528)	(0.1139)	(0.0317)	(0.0217)
6	0.0712	0.1912	0.0508	0.0766***
	(0.0556)	(0.1166)	(0.0335)	(0.0236)
7	0.0534	0.1955*	0.0363	0.1480***
	(0.0564)	(0.1179)	(0.0360)	(0.0271)
8	0.0909	0.1776	0.0402	0.1237***

Table 7 – continued from previous page

Variables	(2001)	(2005)	(2009)	(2015)
	(0.0557)	(0.1194)	(0.0372)	(0.0246)
9	0.1467**	0.1617	0.0252	
	(0.0582)	(0.1205)	(0.0382)	
10	0.1158**	0.2426**	0.0560	
	(0.0591)	(0.1214)	(0.0393)	
11		0.2126**	0.0388	
		(0.1216)	(0.0394)	
12		0.2009**	0.0583	
		(0.1212)	(0.0394)	
13		0.2196*	0.0502	
		(0.1220)	(0.0401)	
14		0.23571*	0.0525	
		(0.1236)	(0.0408)	
15		0.2537**	0.0416	
		(0.1220)	(0.0420)	
16		0.2226*	0.0796*	
		(0.1263)	(0.0419)	
17		0.2647**	0.0788*	
		(0.1250)	(0.0410)	
18		0.2849**	0.0910**	
		(0.1236)	(0.0429)	
19		0.2938**	0.0886**	
		(0.1240)	(0.0445)	
20		0.2274*	0.1225***	
		(0.1330)	(0.0428)	
21		0.2127*	0.0772*	
		(0.1286)	(0.0443)	

Table 7 – continued from previous page

Variables	**(2001)**	**(2005)**	**(2009)**	**(2015)**
22		0.1973	0.0476	
		(0.1253)	(0.0446)	
23		0.2032*	0.0745*	
		(0.1226)	(0.0406)	
24		0.3356***	0.1532***	
		(0.1213)	(0.0400)	
OCCUPY YEAR				
2	-0.0036	0.0175	0.0247	0.0243
	(0.0251)	(0.0282)	(0.0717)	(0.1013)
3	-0.0163	0.0238	0.0769	-0.0362
	(0.0333)	(0.033)	(0.0650)	(0.0955)
4			0.0508	-0.0590
			(0.0632)	(0.0923)
5			0.0449	-0.0118
			(0.0627)	(0.0909)
6	0.1772***		0.0301	-0.0461
	(0.0412)		(0.0622)	(0.0908)
7			0.0390	-0.0484
			(0.0624)	(0.0905)
8		-0.0581*	-0.0089	-0.1071
		(0.0318)	(0.0626)	(0.0912)
9	0.0033	0.0207		
	(0.0203)	(0.0232)		
RACE				
2	0.1089***	0.1065***	0.0628***	0.05214***
	(0.0189)	(0.0198)	(0.0118)	(0.0172)
3	-0.0764	0.1148**	0.0048	-0.0714
	(0.0648)	(0.0551)	(0.0366)	(0.0488)

Table 7 – continued from previous page

Variables	(2001)	(2005)	(2009)	(2015)
4			-0.0450**	-0.0998***
			(0.0218)	(0.0295)
5	-0.1452**	-0.0081	-0.0946	0.0415
	(0.0675)	(0.0297)	(0.0949)	(0.0934)
6			-0.0102	
			(0.0261)	
7	-0.0620**	0.0093	0.0985***	0.0287
	(0.0314)	(0.0292)	(0.0324)	(0.0377)
41	-0.0938**	-0.0615		
	(0.0373)	(0.0419)		
42	-0.0993	-0.1718*		
	(0.1542)	(0.0949)		
IECC				
2B			-0.1938	
			(0.1206)	
3A			0.0128	
			(0.0207)	
3B-4B			-0.2690**	
			(0.1267)	
3C			-0.1779***	
			(0.0533)	
4A			-0.0469	
			(0.0289)	
4C			-0.0150	
			(0.0519)	
5A			-0.1086***	
			(0.0396)	
5B-5C			-0.1829***	

Table 7 – continued from previous page

Variables	(2001)	(2005)	(2009)	(2015)
			(0.0502)	
6A-6B			-0.2070***	
			(0.0485)	
7A-7B-7AK-8AK			-0.2882***	
			(0.0692)	
AIA ZONE				
2			0.0716***	
			(0.0256)	
3			0.1173***	
			(0.0333)	
4			0.1074***	
			(0.0421)	
5			0.1845***	
			(0.0500)	
CLIMATE REGION				
2			0.0542	
			(0.1313)	
3			-0.0665**	
			(0.03012)	
POVERTY				
Poor 100	0.0021	0.0771**	0.0267	
	(0.0357)	(0.0336)	(0.0215)	
Poor 150	0.0057	-0.0239	0.0049	
	(0.0279)	(0.0268)	(0.0173)	
INDICATOR MISSING VARIABLE				
Natural gas	-0.4815***	-0.4643***	-0.4795***	-0.4465***
	0.0173	(0.0183)	(0.0112)	(0.0140)
Heating oil	-0.3582***	-0.4623***	-0.3065***	-0.2924***

Table 7 – continued from previous page

Variables	(2001)	(2005)	(2009)	(2015)
	0.0285	(0.0335)	(0.0226)	(0.0445)
Propane	-0.2880***	-0.2634***	-0.2966***	-0.1860***
	0.0224	(0.0230)	(0.0181)	(0.0226)
Kerosene	-0.0806	-0.0038		
	0.0653	(0.2376)		
OTHER ENERGY USE				
Solar	-0.0077	-0.0124	0.0496	0.1083*
	(0.0570)	(0.1683)	(0.0451)	(0.0566)
Wood	-0.0513***	0.1662***	-0.0476***	-0.0724***
	(0.0173)	(0.0200)	(0.0127)	(0.0180)
Constant	12.222***	11.371***	12.078***	11.159***
	(0.8213)	(0.5888)	(0.4125)	(0.9333)
Observations	3,077	3,002	7,015	3,823
F(74,3002)	91.46	60.14	164.88	144.23
Prob >F	0.0000	0.0000	0.0000	0.0000
R-squared	0.6785	0.6403	0.6830	0.6841
Root MSE	0.3000	0.3233	0.2955	0.2985

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

C Double-selection linear model regression results: Common variables across all of the survey years

Table 8: Double-selection linear model estimates of price elasticities of demand

	2001					2005			
	(1)	**(2)**	**(3)**	**(4)**		**(1)**	**(2)**	**(3)**	**(4)**
Variables	**OLS**	**Plugin**	**CV**	**Adapt**	**Variables**	**OLS**	**Plugin**	**CV**	**Adapt**
Electricity	-0.1587***	-0.1844***	-0.1593***	-0.1457***	Electricity	-0.1678***	-0.2309***	-0.1678***	-0.1641***
	(0.0489)	(0.0420)	(0.0477)	(0.0479)		(0.0494)	(0.0462)	(0.0488)	(0.0481)
Natural gas	-0.3118***	-0.3307***	-0.3140***	-0.3174***	Natural gas	-0.2317***	-0.1993***	-0.2309***	-0.2282***
	(0.0378)	(0.0355)	(0.0372)	(0.0370)		(0.0401)	(0.0381)	(0.0398)	(0.0392)
Heating oil	0.1353	0.0245	0.3014	0.2745	Heating oil	-0.0550	0.2391	-0.1398	-0.0440
	(0.2877)	(0.2532)	(0.3155)	(0.3069)		(0.6946)	(0.7131)	(0.7136)	(0.7116)
Propane	-0.2956	-0.0842	-0.2983***	-0.2763***	Propane	-0.1368	0.0442	-0.1476	-0.1372
	(0.1021)	(0.0857)	(0.0989)	(0.1004)		(0.1035)	(0.0899)	(0.0980)	(0.0980)
MSE	0.0993	0.1519	0.0985	0.0980	MSE	0.1083	0.1310	0.0926	0.0951
R-squared	0.6543	0.4720	0.6560	0.6578	R-squared	0.6004	0.5111	0.6500	0.6453
OS obs.	769	767	763	763	OS obs.	750	746	731	746

	2001					2005			
	(1)	**(2)**	**(3)**	**(4)**		**(1)**	**(2)**	**(3)**	**(4)**
Variables	**OLS**	**Plugin**	**CV**	**Adapt**	**Variables**	**OLS**	**Plugin**	**CV**	**Adapt**
Electricity	-0.1693**	-0.3183***	-0.1693***	-0.1683***	Electricity	-0.2842***	-0.3089***	-0.2848***	-0.2860***
	(0.0288)	(0.0263)	(0.0288)	(0.0288)		(0.0404)	(0.0392)	(0.0403)	(0.0418)
Natural gas	-0.4145***	-0.3113***	-0.4145***	-0.4142***	Natural gas	-0.3139***	-0.3161***	-0.3148***	-0.3150***
	(0.0268)	(0.0257)	(0.0268)	(0.0268)		(0.0282)	(0.0274)	(0.0282)	(0.0282)
Heating oil	-0.7490**	-0.0968	-0.7490	-0.7599**	Heating oil	-0.4900	-0.4125	-0.5121	-0.4466
	(0.3735)	(0.3892)	(0.3735)	(0.3730)		(0.9677)	(0.8840)	(0.9582)	(0.9701)
Propane	-0.2150**	-0.3106***	-0.2150	-0.2132**	Propane	-0.2340**	-0.3295***	-0.2522**	-0.2673**
	(0.0958)	(0.1000)	(0.0958)	(0.0958)		(0.1158)	(0.1247)	(0.1162)	(0.1191)
MSE	0.1013	0.1031	0.0850	0.0855	MSE	0.1053	0.0971	0.0858	0.0863
R-squared	0.6391	0.6290	0.6944	0.6923	R-squared	0.6264	0.6556	0.6956	0.6940
OS obs.	1,754	1,747	1,747	1,747	OS obs.	866	866	866	866

Notes: "MSE" denotes mean-squared-error; "OS obs" denotes out-of-sample observation; "Plugin" denotes a plug-in value method to estimate the lasso tuning parameter; "CV" denotes cross-validation method to estimate the lasso tuning parameter; "Adapt" denotes an adaptive method to estimate the lasso tuning parameter; and "OLS" denotes ordinary least squares. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.