A Top-Down Economic Efficiency Analysis of U.S. Household Energy Consumption

J. Wesley Burnett and Jessica Madariaga***

ABSTRACT

This study analyzes the efficiency of household-level energy consumption using a rich microdata set of homes within the United States. We measure efficiency by extending a cost-minimization model that treats the total amount of energy services produced as latent or unobserved due to technological differences in household consumption. The empirical strategy consists of applying latent class modeling to cost frontier analysis, which helps to identify heterogeneous subsets of units that require the fewest energy resources. Our estimates of efficient units form an empirical cost frontier of best practices within each subset. In order to understand the determinants of household-level energy efficiency, we condition the cost frontier analysis on numerous physical, climate-related, and socio-economic characteristics of the household. We find that state-level energy building code regulations, on average, induce a one-to-four percent marginal increase in household energy consumption.

Keywords: Energy efficiency, energy rebound effect, household energy consumption.

https://doi.org/10.5547/01956574.39.4.jbur

1. INTRODUCTION

Energy efficiency and related demand management policies help mitigate the impacts of climate change by reducing the use of fossil fuels and reducing the energy sector's vulnerabilities to climate change impacts. Over the past forty years, federal and state-level energy efficiency policies (or standards) have been applied to household appliances, the corporate average fuel economy, electric demand-side management programs, weatherization assistance, and building codes. The U.S. residential sector accounts for approximately 21 percent of total primary energy consumption and 20 percent of domestic carbon dioxide emissions (Energy Information Administration (EIA), 2015). Building construction codes and standards regulate the energy efficiency of newly constructed homes or commercial buildings and the energy efficiency requirements specific to renovations, major refurbishments, and the enlargement of buildings. Such codes generally provide minimum building requirements for heating and cooling systems and for any construction or renovations to the housing envelope that leads to energy savings (Aroonruengsawat et al., 2012).

The present study analyzes the determinants of household-level energy use and how efficient each household uses energy compared to a sample of similar homes within the United States. Examining the within-sample efficiency of usage is important as past studies have found that house-

* Corresponding author. Assistant Professor, Department of Economics, College of Charleston, 5 Liberty Street, 413 Beatty Center, Charleston, SC 29401. E-mail: burnettjw@cofc.edu.

** Assistant Professor, Department of Economics, College of Charleston, 5 Liberty Street, 333 Beatty Center, Charleston, SC 29401. E-mail madariagajf@cofc.edu.

holds with similar demographic and economic characteristics, located in the same geographic region, differ considerably in their energy use (Newman and Day, 1975; DOE, 1981a,b; Socolow, 1978).

Given the variation in energy use between households, our paper provides a method to explore the relationship among social, economic, behavioral, and physical factors that determine the pattern and levels of household energy consumption. To this end, we draw on elements from two branches in the economics literature—the energy efficiency gap and the analysis of the determinants of household-level energy efficiency.

This study differs from the recent economic literature associated with analyzing energy efficiency, as our research is not based on a quasi-experimental method or field experiment (Allcott, 2013; Fowlie et al., 2015). A behavioral or experimental framework to examining household energy efficiency is often described as a "bottom-up approach," whereas a neoclassical economic model is often described as a "top-down approach." Bottom up implies that the empiricist does not necessarily make any prior assumptions about a household's behavior or response to a policy, but instead simply observes the response of a treated household in comparison to a similar but non-treated household. Top down, on the other hand, implies that the empiricist formulates prior assumptions about household behavior; such as, the household agent is rational and seeks to minimize costs or produce its energy services efficiently (Orea et al., 2015).

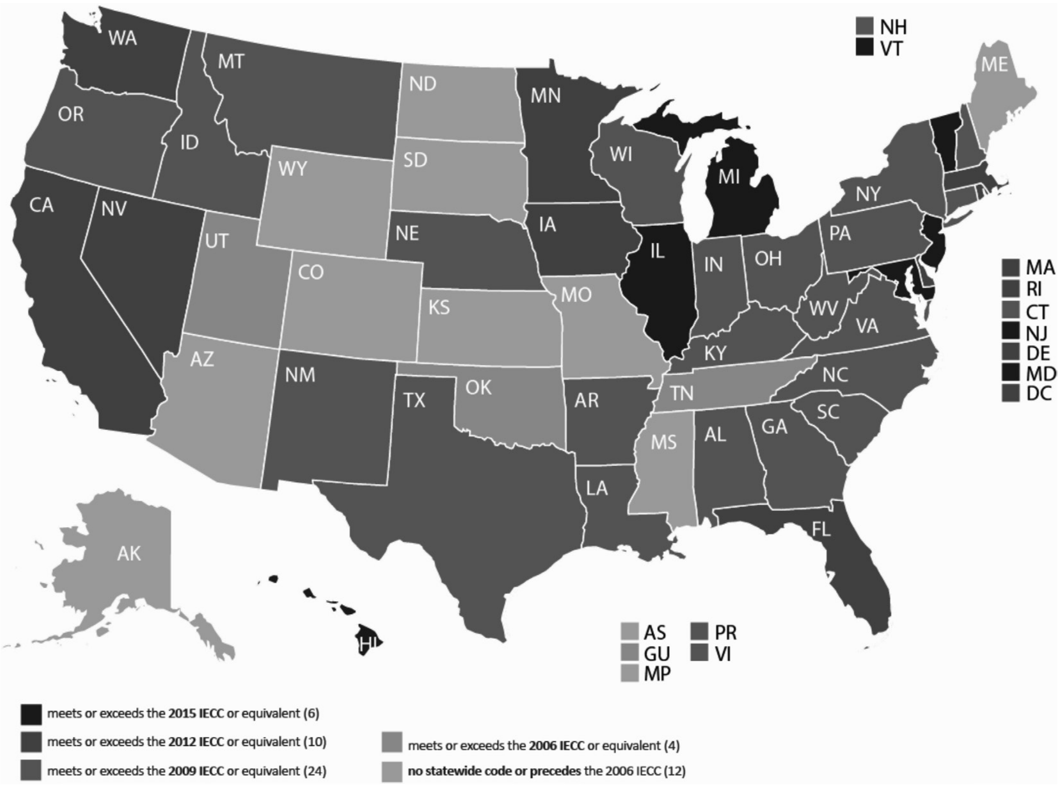
Despite the readers preconceived notions of a behavioral versus a neoclassical approach, each separate framework contains various assumptions, which lead to differing strengths and weaknesses. The current research makes no attempt of arguing in favor of one approach over the other. Instead, we view the separate frameworks as the flip side of the same coin—in terms of analyzing the behavioral response associated with household-level energy efficiency gains—and we proceed by analyzing this phenomenon using a neoclassical approach.

Relative to our neoclassical approach, this study offers five unique contributions to the literature. One, we use a unique data set obtained from the 2009 U.S. EIA's Residential Energy Consumption Survey. The survey contains responses from over 12,000 households across the U.S. providing an incredibly rich cross section of disaggregated data. Two, we develop a theoretical model that demonstrates that a household's energy consumption is affected by the types of energy technologies used within the household. However, the households' energy technologies are not directly observed within the available data, so we use a data-driven method to identify subsets or classes of housing that consume similar levels of energy, arguably based on similar types of household energy technologies. We demonstrate that by dividing the sample into classes, the model offers more accurate measures of efficiency of energy consumption (within each household). Four, the modeling approach allows us to estimate a household-by-household efficiency index of consumption. Five, we condition the cost frontier analysis on a set of other covariates that would potentially affect a household's energy efficiency of consumption, including geographic location, climate zones, and several other household characteristics such as home size and residential makeup.

Our findings imply that the sample contains two unique technological classes (or groups) and that the estimated efficiency indexes for each class is statistically different from the other class and the overall sample. Based on these insights, we proceed by examining the cost frontier estimates by assuming heteroskedasticity within the inefficiency term, and then explore the marginal effects of residential energy building code policies on household-level (in)efficiency of energy consumption.

Our estimates suggest a relatively small but highly statistically significant (partial) energy rebound effect, where the energy code policies, on average, lead to an increase in residential energy consumption. These findings are similar to that of Levinson (2016), who found *no* evidence that

Figure 1: State Residential Energy Code Status, 2016



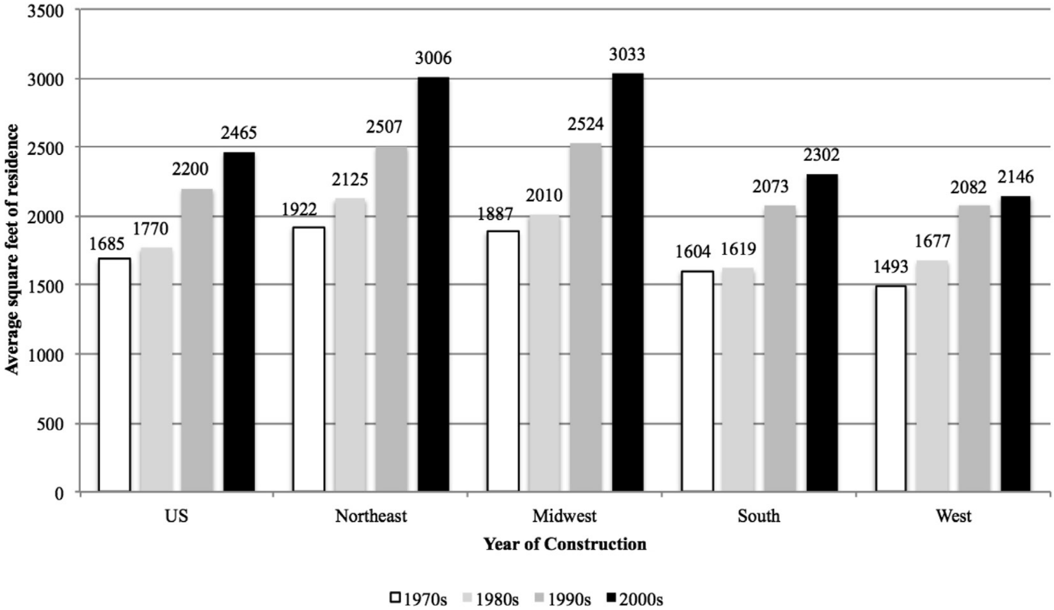
Source: BCAP (2016)

California homes, constructed since the State instituted its building energy codes, use less electricity today than homes built before the codes came into effect. Our findings offer insights to future building energy code policy formulation based on economic analysis, in contrast to engineering analysis, of the efficiency of residential energy consumption.

1.1 Residential energy efficiency policies in the U.S.

Energy efficiency requirements for new buildings, including residential homes, are important because such requirements determine the building sector's energy consumption for far longer than other end-use sectors (Laustsen, 2008). Improvements in a building's energy efficiency are much easier, and more cost effective, to implement in the planning stage; whereas, improvements after the initial construction phase are generally more costly. In 1978, California became the first state in the U.S. to adopt state-level energy requirements in its building codes (EPA, 2015). Today, forty states and the District of Columbia have some form of residential-energy building codes. A heat map of current U.S. states, with residential energy code requirements, is provided in Figure 1. (The term "IECC" in the figure refers to the International Energy Conservation Code, which provides standards and metrics for energy conservation and efficiency code requirements). Such codes arguably offer a range of benefits including lower energy use, reduced energy costs, reduced pollution emissions, stronger local economies, improved energy resource reliability and improved health (EPA, 2015). The Department of Energy (DOE) (2014) estimates that from 1992 to 2012 such codes

Figure 2: Newer Homes Trending Larger in the U.S., 1970–2010



Source: EIA (2012)

led to approximately 4.2 quads (42[15] British thermal units (BTUs)) in cumulative energy savings and $44 billion in cost savings to consumers. The same DOE report suggests that current building energy codes could potentially save a total of 46 quads (460[15] BTUs) of energy through 2040, which is approximately equal to one year's consumption of energy within the U.S. residential sector.

While the U.S. residential housing sector has become more energy efficient over the past two or three decades, the EIA (2012) also reports that the overall size of new U.S. housing builds has been increasing over the past couple of decades. Figure 2 demonstrates the upward trend in the average size of a newly built home (measured in square feet) in the U.S. from 1970 through 2010. A larger space within a home often implies greater demands for heating and cooling needs, however, larger homes often tend towards adopting additional energy efficiency features. Given all these factors, whether there is a net increase or decrease in energy consumption is an empirical question.

2. RELATED WORK

2.1 The energy efficiency gap and energy efficiency analysis

Despite the purported savings associated with energy efficiency investments, McKinsey and Company (2007) claims that the U.S. has over $100 billion in energy-saving opportunities that have been left unrealized. In other words, U.S. households have figuratively "left free money laying on the ground" by not realizing considerable energy and financial-saving opportunities associated with the adoption of energy efficiency retrofits and/or technologies.

Are there indeed significant windfalls of unrealized financial and energy savings for U.S. households or are McKinsey and Co.'s projected savings overstated? In general, the economics literature seems to indicate that there are savings to be gained, but the savings are less than estimates based on engineering analysis. Herein lies the problem—energy efficiency estimates, including those offered by McKinsey and Co., are generally based on engineering analysis, not economic analysis. Joskow and Marron (1992) describe many factors that contribute to the overstatement of program effectiveness—namely, rebound effects (further explained below) and other confounding factors. In the context of building code policies, the engineering analysis generally presumes that energy building codes (or standards) will be enforced, the calculated engineering savings will be realized, and there is no behavioral response (Levinson, 2016).

There are two interrelated branches in the economics literature that address the gap between engineering and economic cost analysis. The first branch explores the "energy paradox" or the "energy efficiency gap." This framework tries to address what could be causing the paradox, where energy-efficient technologies offer a reduction in financial costs and a reduction in environmental damages associated with energy use. Yet, these same technologies are not adopted by households to the degree that appears justified. There have been a number of different explanations offered including market failure, information asymmetry, myopia, cognitive limitations, loss aversion, and systematic biased beliefs (Gerarden et al., 2015; Allcott, 2013; Anderson and Newell, 2004; Davis et al., 2014; Gillingham et al., 2012; Greene et al., 2013; Jaffe and Stavins, 1994; Newell and Siikamäki, 2014; Sallee, 2013).

The second branch in the literature examines the economic determinants of energy efficiency in the context of relating building codes to energy savings. For example, Aroonruengsawat et al. (2012) estimate that U.S. energy building codes saved, on average, two to five percent in residential electricity consumption in 2006. Jacobsen and Kotchen (2013) also find that the increased stringency in Florida's energy code in 2002 led to a four percent decrease in household electricity consumption. While the above studies highlight the relationship between building codes and energy costs, questions remain regarding the economic magnitude of real world returns on energy efficient investment. Based on a randomized controlled evaluation of the Federal Weatherization Assistance Program in Michigan, Fowlie et al. (2015) find that although participating households' energy consumption declined by 10 to 20 percent on average, the savings were approximately 60 percent less than the savings predicted by the engineering models. Another study shows that adding additional insulation in a home's attic provides household energy saving of 10 percent, on average, but the return on investment is smaller than the engineering estimates (Metcalf and Hassett, 1999).

Despite fairly progressive building energy codes in California, Chong (2012) finds that newer constructed homes, on average, exhibit a higher level of response to changes in higher temperatures than a comparable sample of older homes in Southern California. In other words, homeowners of newer homes turned down the thermostat in response to periods of higher outdoor temperatures by a greater percentage than did older homes, leading to higher energy consumption. Similarly, Levinson (2016) argues that newly constructed or renovated homes, built after California's building codes were enacted, use no less energy than homes built before the codes were established. Given the claim by the California Energy Commission (2013) that it has saved the State's residents more than $74 billion in reduced electricity bills since 1997, the findings of Levinson and Chong are especially thought provoking and highlight the importance in distinguishing between costs calculated through engineering and economic analysis.

Our results confirm the findings of earlier studies, such as Baxter et al. (1986), Chong (2012), and Levinson (2016), that household demographic composition and building vintage are important determinants of household-level energy consumption and efficiency. In addition to demographics and building characteristics, we control for climate-related factors that affect household demand for energy services.

Similar to Baxter et al. (1986), this study contributes to the literature by examining a theoretical cost frontier of energy services to gain insights into how households consume energy. To our knowledge, the works by Baxter et al. (1986) (followed more recently by Orea et al. (2015) and Filippini and Hunt (2012)) are the only studies that have used production theory (and stochastic frontier analysis) to better understand household-level energy usage. Our current research improves upon the original work by using significantly more observations and analyzing the efficiency of household energy consumption with a latent class, cost frontier analysis. Our approach differs slightly from recent works including Chong (2012) and Levinson (2016), who examined household electricity consumption, after controlling for the vintage of the home and other characteristics. Our particular methodology, examining the overall technical (in)efficiency of energy usage, is simply a different approach to better understand the determinants of within-sample energy efficiency.

2.2 Households as productive units

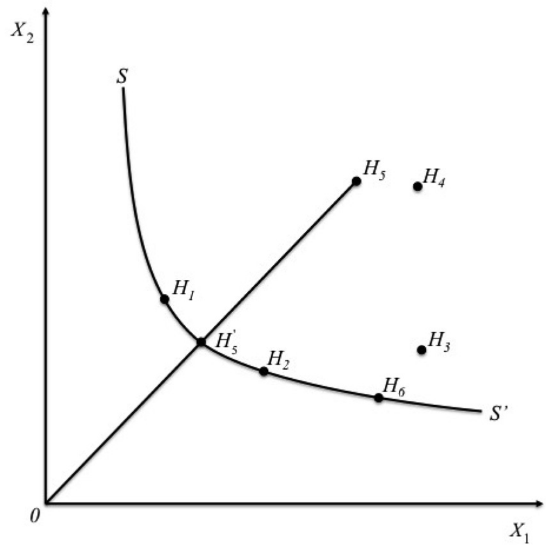
Households, like firms, are economic agents in which members of the household carry out the production of goods and services (Becker, 1981; Ironmonger, 2000; Pollak, 2003). Goods and services produced by households include accommodations, meals, childcare, etc. And similar to firms, households use labor and (human) capital as its factors of production. Becker (1981) established two foundational (economic) assumptions regarding household production—behavior maximization and market equilibrium.

We assume that a household's energy consumption is treated as a derived demand (i.e., electricity is not demanded in-and-of-itself, but rather for the services it provides), and that energy is an input in the provision of a range of household services. For example, electricity or natural gas can be used to provide indoor heating and cooling, hot water heating, and lighting, among other uses. We also maintain one of the foundational assumptions established by Becker (1981), in which households seek to maximize utility through their consumption of energy services. Since energy is a derived demand, households do not necessarily seek to consume the most energy possible in order to maximize utility, but rather, households demand for energy is constrained by the costs of providing such services (Scott, 1980). If a household's utility is derived from its accumulation of lifetime wealth, then this notion of cost minimization is compatible with a household's utility derived from wealth maximization (or accumulation).

We treat the household's dwelling unit (and its inherent energy using equipment) as the fixed factor(s) of production, and the household's input of fuels (electricity or natural gas) are treated as variable factors. The household is capable of reducing its energy consumption on the intensive and extensive margin. Along the intensive margin, the household can reduce its consumption by simply modifying its monthly energy usage in the short run, such as adjusting the thermostat for indoor heating or cooling. Along the extensive margin, the households can reduce consumption by investing in a more energy efficient technology (e.g., purchasing a tankless water heater) or by retrofitting its home with more energy efficient windows or insulation, which arguably reduces fuel consumed per each unit of service output in the long run.

The firm's (or household's) production process is defined as technically efficient if it is producing the maximum output for a given set of inputs. However, the same definition of efficiency holds for a production process that uses the minimum level of inputs to produce a given set of outputs. Technical efficiency is related to the concept of productive (or allocative) efficiency, where firms (or households) minimize costs for a given level of output (Varian, 1992).

Figure 3: An efficiency frontier and the relative index of efficiency



Source: This figure was originally produced by Baxter et al. (1986), and we reproduced it here for illustration.

The productive efficiency of a firm can be measured by the ratio of its output to inputs (Farrell, 1957; Lovell, 1993). As an illustration, consider the efficiency frontier diagram in Figure 3 based on the original work of Baxter et al. (1986) and Farrell (1957). The measure on the vertical axis is a hypothetical amount of one input (such as natural gas). Similarly, the measure on the horizontal axis is a hypothetical amount of the other input (such as electricity). (Figure 3 is purely for illustrative purposes and is not based explicitly on the data observed within the current study). In Figure 3, the efficiency frontier measures the relative amount of a particular energy service provision, such as home heating and cooling, given the limited set of inputs. Stated differently, the frontier explores the least inputs required to consume a comparable level or unit of said energy service. The former is concerned with the absolute amount of energy service provision, whereas the latter is concerned with the efficiency of consumption (or production) of the same service provision. (For the time being we will abstract away from "allocative inefficiency," in which a household could potentially save costs by reallocating its inputs from one input to the other; however, we will return to the discussion of allocative inefficiency later in the study).

Following this hypothetical example, the household energy service provision is for home heating and cooling that could require two inputs (natural gas and electricity) and results in one output (temperature control). Figure 3 illustrates the service provision for six hypothetical households (H_1–H_6) distributed as input-per-unit output space. The physical location of each household (in the figure) represents its consumption of inputs one (X_1) and two (X_2). The isoquant, SS', is the inner boundary of the input set, reflecting the minimum input combinations that may be used to produce a given output vector. Therefore, the curve, which is convex to the origin, illustrates the best practice (or cost-minimizing) frontier (i.e., the least amount of inputs consumed per unit of output produced) relative to the other households (Coelli et al., 2003). In this particular example, X_1 could represent the electricity consumed per kilowatt hour and X_2 could present the amount of natural gas consumed per therm. As demonstrated in the figure, households one (H_1) and two (H_2) are more efficient than household five (H_5) as both use fewer inputs per unit of output. Similarly, households two (H_2) and six (H_6) are more efficient than household three (H_3). Conversely, households H_3, H_4, and H_5 are inefficient relative to households H_1, H_2, and H_6, as the latter three demonstrate that it is possible to produce the same level of output with fewer input resources. The line segment (or curve) connecting H_1, H_2, and H_6 forms the efficiency frontier, which represents a best practice (or cost-minimization) pattern for this particular service provision.

Based on this diagrammatic example, we can measure the efficiency level of each of the inefficient units (H_3, H_4, and H_5) by holding constant the mix of inputs used by these units. Measuring efficiency in this regard is equivalent to drawing a line from the origin and through the frontier to one of the inefficient units (e.g., unit H_5 as illustrated in Figure 3). The point marked as H_5' represents the equi-proportional reduction in inputs, per unit of output, which would bring the housing unit H_5 to the best practices frontier. An index for the relative efficiency of household H_5 can be calculated as the ratio of the length of line segment OH_5' to the length of line segment OH_5 (Baxter et al., 1986; Scott, 1980). The resulting index then represents the efficient proportion of inputs needed for a technically efficient service provision (Farrell, 1957). This ratio, OH_5' / OH_5, ranges between zero and one where a value of one indicates that the service provision is on the best practices frontier. In general, the larger the value, the more productive the household service provision is compared to other households with the same energy input mix. In this particular example, we use only two input dimensions and one output dimension, but the analysis can easily be expanded to multiple inputs and the interpretation would remain the same—i.e., the efficiency index will still vary between values of zero and one (Farrell, 1957), provided we hold the other factor inputs fixed.

It is worth pointing out that this study does not seek to measure a input distance function, as implied by Figure 3 and the example above. Rather, we are interested in measuring the determinants of technical (in)efficiency—i.e., household-level behavior that deviates from the theoretical cost-minimizing frontier—in a stochastic cost frontier analysis. The figure and example above are purely for illustrative purposes.

2.3 Theoretical model of input-oriented technical inefficiency

We assume that the objective of the household is to produce a given level of energy services with the minimum possible costs of inputs. Further, we assume that the household is technically inefficient in its use of energy—that is, it either produces less than the maximum possible output or it uses more inputs than is necessary to produce a given level of output. Therefore, the model we present below is an input-oriented measure of technical inefficiency.

Given these assumptions, the cost minimization problem for a representative household is given by:

$$\min_{\mathbf{x}} \mathbf{w}'\mathbf{x} \quad \text{s.t.} \quad q^* = f(\mathbf{x} \cdot e^{-\eta}), \tag{1}$$

where $\mathbf{w}$ denotes a vector of input prices, $\mathbf{x}$ denotes a vector of energy inputs, q^* denotes the latent (unobserved) but true consumption (or derived demand) of energy service (outputs), and η denotes a input-oriented technical inefficiency term (parameter). The technical inefficiency term is assumed to be nonnegative, $\eta \geq 0$, and it measures the percentage by which all the inputs are overused in producing some level of output q^*. We further assume that the production function is homogeneous of degree θ in the energy inputs $\mathbf{x}$, such that if each of the inputs is multiplied by a positive scalar factor, c, then it would yield the following:

$$q^* = f(c \cdot \mathbf{x} \cdot e^{-\eta}) = f(\mathbf{x} \cdot e^{-\eta}) \cdot c^{\theta}. \tag{2}$$

The first-order conditions (Kumbhakar et al., 2015) for the cost minimization problem are given by

$$\frac{f_j(\mathbf{x}\cdot e^{-\eta})}{f_1(\mathbf{x}\cdot e^{-\eta})}=\frac{w_j}{w_1},\quad j=2,\ldots,J, \tag{3}$$

where f_j denotes the derivative of the production function with respect to the j^{th} energy input; w_j denotes the input price associated with the j^{th} energy input; and, J denotes the total energy inputs in the house's provision of energy services.

Given the assumptions outlined above, there are $J-1$ first-order conditions, from which we can solve for the conditional-factor (or input) demand functions (Kumbhakar et al., 2015):

$$x_j\cdot e^{-\eta}=\psi_j(\mathbf{w},q^*),\quad j=1,\ldots,J, \tag{4}$$

where ψ_j denotes the input-demand function, associated with the j^{th} energy input as a function of the input prices and the latent consumption of energy services. Based on the input demand functions, we can derive the cost function of energy services:

$$C(\mathbf{w},q^*)=\sum_j w_j\cdot x_j\cdot e^{-\eta}. \tag{5}$$

Equation (5) is a theoretical, *frontier* cost function, as it provides the minimum costs associated with the vector of input prices and the true level of energy demand within the home (Kumbhakar et al., 2015). We can apply Shephard's Lemma to the cost function to derive

$$\frac{\partial \ln C(\cdot)}{\partial \ln w_j}=\frac{w_j\cdot x_j\cdot e^{-\eta}}{C(\cdot)}=\frac{w_j\cdot x_j}{\mathbf{w}'\mathbf{x}}\equiv S_j, \tag{6}$$

where S_j denotes the cost share of the j^{th} energy resource.

Given the assumptions of cost minimization, we can write the observed, household-level of energy costs as a function of the theoretical, cost-minimizing level of energy costs (Kumbhakar et al., 2015):

$$C^a=\sum_j w_j\cdot x_j=C(\cdot)\cdot\exp(\eta), \tag{7}$$

where C^a denotes the actual or observed total costs. Transforming equation (7) by taking its natural logarithm allows the equation to be re-expressed as

$$\ln C^a=\ln C(\mathbf{w},q^*)+\eta. \tag{8}$$

Based on this equation, we can then derive the *efficiency* (of some representative household) measure of consumption as

$$\exp(-\eta)=\frac{C(\cdot)}{C^a}, \tag{9}$$

where the ratio is bounded between 0 and 1, and the estimation of the efficiency index is numerically guaranteed by imposing $\eta\geq 0$ (Kumbhakar et al., 2015). Conversely, we could measure the *inefficiency* of consumption (of a representative household) as:

$$\eta=\ln C^a-\ln C(\mathbf{w},q^*). \tag{10}$$

The empirical counterpart to equation (9) is $E[exp(-\eta)|v]$ (where v is a random noise term and $E[\cdot]$ is an expectation operator)—this is discussed further below. By converting the estimated technical inefficiency term, $-\eta$, into exponential form, the conversion yields a positive estimate bounded on the interval [0,1). The positive estimate is often referred to as an (technical) efficiency index of household consumption. An estimated index value of one would indicate that a household's actual consumption is equal to the theoretical cost-minimizing level of consumption; or diagrammatically, the household's observed level of consumption would be on the cost-minimizing frontier (as in Figure 3). Generally speaking, a household's consumption is more efficient the closer the estimate is to one and less efficient the closer the estimate is to zero.

3. METHODOLOGICAL APPROACH

Related to Baxter et al. (1986) and consistent with the theoretical approach detailed in Section 2, we model the cost minimization for a household's costs of energy services. Our model provides a measure of technical efficiency by evaluating the efficiency of the physical transformation of resource inputs (such as electricity consumption) to service outputs (such as home heating or cooling). Input prices are measured in dollars per unit of measurement; e.g., the price of electricity is measured as $US per kilowatt hour and the price of natural gas is measured as $US per therm. The total household consumption of energy is measured in British thermal units (BTUs), which provides a common metric for the different types of energy resources consumed. A British thermal unit is a measure of the energy required to raise the temperature of one pound of water by 1° Fahrenheit—it is a general measure of the "heating value" of a particular fuel (Hinrichs and Kleinbach, 2006).

3.1 Empirical model: Production possibility frontier analysis

Historically, there have been two principle methods to estimate a production or efficiency frontier—through a stochastic frontier analysis (SFA) or a data envelopment analysis (DEA). The stochastic frontier analysis accounts for failures (technical inefficiency) in production optimization (Meeusen and van Den Broeck, 1977; Aigner et al., 1977), while the DEA approach is a nonparametric specification of the production frontier and is deterministic. Because the DEA approach is deterministic—meaning it does not contain a stochastic or random component—it is generally not suitable for a statistical analysis of productive efficiency. However, the SFA approach is more appropriate for use in a statistical or econometric regression analysis.

We estimate the model specified in equation (8) by adding a random noise term v to the right-hand side of the equation and defining a particular distribution for both the noise term and the technical inefficiency term (details are discussed in subsection 3.2). A non-restrictive way to define the cost function outlined above is by using a transcendental logarithm or "translog" specification (Christensen and Greene, 1976) on $\ln C(\cdot)$ as

$$\begin{aligned}\ln C_i^a &= \ln C(q_i^*, \mathbf{w_i}) + v_i + \eta_i \\ &= \beta_0 + \sum_j \beta_j \ln w_{j,i} + \beta_{q^*} \ln q_i^* + \frac{1}{2}\sum_j\sum_k \beta_{jk} \ln w_{j,i} \ln w_{k,i} + \frac{1}{2}\beta_{q^*q^*} \ln q_i^* \ln q_i^* \\ &+ \sum_j \beta_{jq^*} \ln w_{j,i} \ln q_i^* + v_i + \eta_i, \end{aligned} \tag{11}$$

where we have added a subscript i to the above equation to denote a unit of observation for the i^{th} household. As previously mentioned, we do not directly observe the true demand for energy services, q^*; nevertheless, we use the term within the translog specification above for illustrative purposes. Looking ahead, we potentially solve this problem through the latent class analysis, where homes with similar exogenous attributes will arguably use similar types of energy technologies allowing for a more accurate measure of $q_i^* = q_i^a |_g$ (based on the actual observed level of demand q_i^a) for the i^{th} house in the g^{th} class of similar homes. By definition, the cost function above is homogenous of degree one in the input prices (Varian, 1992), so we can specify the following restrictions to satisfy that condition:

$$\sum_j \beta_j = 1, \quad \sum_j \beta_{jk} = 0\,\forall k, \quad \sum_j \beta_{jq^*} = 0. \tag{12}$$

The flexibility of the translog specification in (11) allows us to test the below restrictions within the model:

Hypothesis one: No restrictions on the parameters;
Hypothesis two: $q_{ig}^* = q_i^*, \forall i$ (No differences in technical efficiencies between classes);
Hypothesis three: $\beta_{jk} = 0$ (Cobb-Douglas production specification);
Hypothesis four: $1 - \dfrac{\partial \ln(C)}{\partial \ln(q^*)} = 1 - \beta_{q^*} = 0$ (Constant returns to scale technology).

3.2 Latent class analysis

Unobserved heterogeneity, where differences across observations are not reflected within the data, is problematic within an empirical framework and the solution to dealing with the heterogeneity depends on its relationship with the explanatory variables. If the differences in the unobserved heterogeneity are not correlated with the explanatory variables, the problem can be accommodated through the error term. However, if the differences in the unobserved heterogeneity are correlated with the explanatory variables, the resulting estimated parameters will be biased (Griliches, 1957). To overcome this problem, it is often necessary to use models that are able to estimate different parameters for each group. If the number of heterogeneous groups are discrete and we can potentially estimate as many technologies as there are groups, then a cluster analysis or latent class model can help alleviate the problem (Alvarez and del Corral, 2010).

A cluster analysis is a statistical technique that stratifies the sample into several groups based on similarities among the members of each group (Aldenderfer and Blashfield, 1984). Using this approach, the cluster analysis would presumably be carried out in an initial step, followed by the SFA approach in a separate step. A latent class model, on the other hand, assumes a finite number of classes within the underlying data and can be carried out in conjunction with SFA in one single step, thereby reducing the potential bias of a two-step estimation procedure (Beard et al., 1991; Orea and Kumbhakar, 2004; Greene, 2005; Alvarez and del Corral, 2010; Brown et al., 2014).

Following equation (8), and adding a noise term v to the right-hand side, we can specify the latent class approach as

$$\ln C_i^a = \ln C^*(q_i^a, \mathbf{w}_i)\Big|_g + \varepsilon_i\Big|_g, \tag{13}$$

where ε_i is a composite error defined as $\varepsilon_i = v_i - \eta_i$. Otherwise, the superscript a denotes the observed total costs of energy services and the observed level of total energy demand; and, the subscript i denotes a unit of observation on the i^{th} house. The subscript g denotes a latent class index ($g = 1,2,...,G$). In turn, the vertical bars denote a different model for each class, g.

Assuming that v is normally distributed ($v \sim N(0,\sigma_v^2)$) and η follows a half-normal distribution ($\eta \sim N^+(0,\sigma_\eta^2)$), the likelihood function (LF) for each house i within class g is (Greene, 2005; Kumbhakar et al., 2015):

$$LF_{ig} = -\ln\left(\frac{1}{2}\right) - \frac{1}{2}\ln(\sigma_g^2) + \ln\phi\left(\frac{-\varepsilon_i\,|_g}{\sigma_g}\right) + \ln\Phi\left(\frac{\mu_*}{\sigma_*}\right), \tag{14}$$

where $\varepsilon_i = \ln C_i^a - \ln C^*(q_i^a, \mathbf{w}_i)$, $\sigma_g^2 = (\sigma_{\eta g}^2 + \sigma_{vg}^2)$, $\mu_* = (\sigma_{\eta g}^2 \cdot \varepsilon_i)/\sigma_g^2$, and $\sigma_*^2 = (\sigma_{\eta g}^2 \cdot \sigma_{vg}^2)/\sigma_g^2$. Further, the terms ϕ and Φ denote the standard normal density and cumulative distribution function.

The specification in (14) implicitly assumes that the model is composed of homoskedastic errors. However, the estimation of stochastic frontier models in the presence of heteroskedasticity, if not accounted for, can yield biased parameter estimates (Caudill et al., 1995; Wang and Schmidt, 2002). Therefore, we chose to forego the classic stochastic frontier estimates, which assume homoskedastic errors, and opted instead for a specification that controls for heteroskedastic errors.

Thus, before proceeding it would be informative to offer just a bit more notation for the sake of completeness. Following Caudill et al. (1995) and Kumbhakar et al. (2015), heteroskedasticity can be parameterized by a set of observed variables and associated parameters. Kumbhakar et al. (2015) offer the following parameterization (which we followed in our empirical analysis herein):

$$\sigma_{\eta,i}^2 = exp(\mathbf{z}_{\eta,i}' \cdot \mathbf{w}_\eta), \tag{15}$$

$$\sigma_{v,i}^2 = exp(\mathbf{z}_{v,i}' \cdot \mathbf{w}_v), \tag{16}$$

where the vectors $\mathbf{z}_{\eta,i}$ and $\mathbf{z}_{v,i}$ may or may not contain the same vector of observed variables, and they may also contain all or part of the explanatory variables in $\mathbf{x}_i$. The term $\mathbf{w}_\eta$ denotes a vector of parameters associated with the observed variables. We modeled heterogeneity by specifying an indicator variable for states that have rigorous residential building codes (otherwise, the vector of observables within $\mathbf{z}_\eta$ contains a constant term as well)—this is discussed in further detail below. Moreover, we assume σ_v^2, the variance of the noise term, is constant.

Given these assumptions, the likelihood function for house i is obtained as a weighted average of its likelihood function for each class g, using the prior probabilities of class g membership (Greene, 2005; Alvarez and del Corral, 2010):

$$LF_i = \sum_{g=1}^{G} P_{ig} \cdot LF_{ig}, \tag{17}$$

where G denotes the total number of estimated classes. We parameterize the probability P_{ig}, of the i^{th} house in the g^{th} class, as a multinomial logistic (MNL) (Orea and Kumbhakar, 2004; Alvarez and del Corral, 2010):

$$P_{ig} = \frac{\exp(\delta_g \cdot \mathbf{z}_i)}{\sum_{g=1}^{G} \exp(\delta_g \cdot \mathbf{z}_i)}, \tag{18}$$

where $\mathbf{z}_i$ is a vector of exogenous determinants that are used to stratify the sample, and δ_g is a vector of parameters to be estimated. One class is chosen as the reference for the other classes in the multinomial logistic function.

Following Greene (2002) and Orea and Kumbhakar (2004), we can define the posterior class probability as

$$P(g\,|\,i) = \frac{LF_i \cdot P_{ig}}{\sum_{g=1}^{G} LF_i \cdot P_{ig}}. \tag{19}$$

The posterior class probabilities depend not only on the estimated prior class probabilities (based on the estimated parameters, δ_g, from the MNL model), but also on the parameters from the cost frontier model. Therefore, the latent class model stratifies the sample into several groups even though actual sample-separating information is not available (Orea and Kumbhakar, 2004). As illustrated in the above expression, the latent class structure uses the goodness of fit of each estimated frontier, in conjunction with the prior class probabilities, to identify groups or classes of households.

The log-likelihood function was maximized with respect to the parameter set $\theta_g = (\beta_g, \delta_g, \sigma^2_{ng}, \sigma^2_{vg})$ using Stata 14.1 with our own programmed code, which borrows heavily on the Stata programs provided by Kumbhakar et al. (2016).

An important issue in estimating these types of models is how to determine the number of unobserved classes. Greene (2005) suggests to pre-specify a beginning value of G^*, which is at least as large as the true G, and test downward with subsequent smaller classes nested within the initial specified G^* number of classes. Each subsequent specification can be tested against the pre-specified number of classes by utilizing likelihood ratio tests and/or Information Criteria post-estimation diagnostics. Alvarez and del Corral (2010) suggest the following diagnostics:

$$SBIC = -2 \cdot log\,LF(G) + m \cdot log\,n,$$

$$AIC = -2 \cdot log\,LF(G) + 2 \cdot m,$$

$$CAIC = -2 \cdot log\,LF(G) + 2 \cdot (1 + log \cdot n),$$

$$HQIC = -2 \cdot log\,LF(G) + 2 \cdot log \cdot log\,m,$$

where SBIC denotes the modified Schwartz Bayesian Information Criterion, AIC is the Akaike Information Criterion, CAIC is the consistent Akaike Information Criterion (Bozdogan, 1987), and HQIC is the Hannan and Quinn (1979) Information Criterion. The term $LF(G)$ denotes the value of the likelihood function for G groups (or classes); m denotes the total number of parameters used in the model; and, n refers to the number of observations within the sample.

The AIC statistic has been criticized for its tendency to favor large models (Brown et al., 2014); therefore, Bozdogan (1987) developed the CAIC statistic, which in theory provides a consistent measure of the AIC statistic despite the size of the model. Through a series of Monte Carlo experiments, Bozdogan (1987) demonstrated that the CAIC performed better than the traditional AIC in terms of providing consistent estimates of the statistic. Despite its inconsistency, we report the AIC statistic as baseline against the other criterion statistic estimates. The favored model is that which provides the lowest values of the four metrics (Alvarez and del Corral, 2010).

Once the class assignment is completed, the efficiency index for each household is computed using the frontier assigned for that class as its *reference* technology. This method, however, ignores all other class probabilities even if the posterior class probabilities are non-zero (Orea and Kumbhakar, 2004). Thus, we can potentially avoid arbitrary weighting, and selection of the so-called reference technology, by using the following method to calculate household efficiency:

$$\ln EF_i = \sum_{g=1}^{G} P(j\,|\,i) \cdot \ln EF_i(g), \tag{20}$$

where $P(j|i)$ is the posterior probability, (19), in the g^{th} class for a given household i, and EF_i denotes the efficiency index measure relative to the reference technology in class g (the EF_i term is based on equation (9) above). This strategy, suggested by Greene (2002), takes into account technologies from every class. Our empirical estimates of the efficiency index are based on what Alvarez et al. (2006) describe as the "RSCFG specification" (Reifschneider and Stevenson, 1991; Caudill and Ford, 1993; Caudill et al., 1995). To calculate the "adjusted" efficiency index (based on partial rebound effect) in Table 9, we utilized equation (A7) outlined in Orea et al. (2015, Appendix A).

4. DATA AND DATA ANALYSIS

The data used in this study are from the 2009 Residential Energy Consumption Survey (RECS), which was conducted by the EIA (2012). The aim of the survey was to measure annual energy consumption from a large national sample of households, as well as to collect data on household unit characteristics associated with energy consumption. The RECS data set initially contained over 12,050 observations; however, after cleaning the survey data, this study resulted in approximately 12,007 usable household observations. The data set includes household-level expenditures and use of various energy resources, as well as numerous physical, social, and economic characteristics of the household units. The descriptive statistics are displayed in Tables 1–4.

Two sets of variables are needed to estimate the model introduced in Section 3: the variables for the stochastic cost frontier model (i.e., the total cost and amount of energy consumption, the energy input prices, and other controls); and the variables to help determine prior class probabilities.

To calibrate the prior class probabilities, we examine the intensity of household energy consumption, where we define intensity as the ratio of total energy consumption to the total amount of square feet within the house. We divide the overall sample into initial (pre-specified) class specifications based on the quantiles of energy intensity. For example, in a two-class model we divide the overall sample into two quantiles—a relatively low-intense sample and a relatively high-intense sample. We follow this procedure up to a total of eight separate class stratifications. Given that the log-likelihood function does not converge for the seven-class (nor the eight-class) model, we limit the analysis to six potential classes. The variables used in the prior class probabilities included: a set of ten dummy variables representing the Census division where the house was located; a set of five dummies representing pre-specified climate regions (cold, hot-dry, hot-humid, mixed-humid, and marine); a set of five (pre-specified) American Institute of Architects (AIA) climate zones (based on average temperatures from 1981–2010); and, two additional variables representing the number of cooling and heating degree days for each observation.

In order to estimate the latent class model, we first address the problem of determining the total number of classes. Although the Akaike Information Criterion (AIC) and (Schwartz) Bayes Information Criterion (BIC) are the most widely used metrics to identify the number of classes in standard latent class models (Orea and Kumbhakar, 2004), we also included the consistent AIC (Bozdogan, 1987) and the Hannan and Quinn Information Criterion (HQIC) as additional metrics for evaluation. All four statistics provide an overall measure on the goodness-of-fit of the underlying data to the corresponding model, and each Criterion puts a penalty on the number of parameters used in the model. Based on these metrics, we compare the goodness-of-fit of the six separate pre-specified classes. There is no general consensus on the optimal criterion within the latent class modelling literature as the individual criteria metrics are derived from different principles and as a results have differing properties (Brown et al., 2014). However, as Table 5 highlights, all four crite-

Table 1: Descriptive statistics—energy usage characteristics

Variable	Obs	Mean	Std Dev	Min	Max
Annual energy consumption data					
Total costs of energy services ($US)	12,007	2,038	1,174	75	32,012
Total output of energy services (million BTUs)	12,007	90.18	54.44	1.89	1,096.08
Electricity prices ($US per kilowatt-hour)	12,007	0.13	0.05	0.02	1.73
Natural gas prices ($US per therm)	12,007	1.26	0.35	0	3.5
Annual expenditures on energy resources					
Electricity ($US)	12,007	1,353	905	50	19,040
Natural gas ($US)	12,007	492	578	0	6,355
Energy building code regulations					
Meets or exceeds 2012 IECC standards	12,007	0.39	0.49	0	1
Share of annual energy budget					
Electricity	12,007	0.68	0.24	0.03	1
Natural gas	12,007	0.25	0.24	0	0.92
Cooling and heating degree days					
HDD	12,007	4,136	2,260	0	13,346
CDD	12,007	1,443	1,021	0	5,357

Table 2: Descriptive statistics—climate region and locational characteristics

Variable	Obs	Mean	Std Dev	Min	Max
Building America climate region					
Very Cold/Cold	12,007	0.33	0.47	0	1
Hot-Dry/Mixed-Dry	12,007	0.14	0.35	0	1
Hot-Humid	12,007	0.18	0.38	0	1
Mixed-Humid	12,007	0.29	0.45	0	1
Marine	12,007	0.06	0.23	0	1
Census division					
New England	12,007	0.08	0.27	0	1
Middle Atlantic	12,007	0.11	0.31	0	1
East North Central	12,007	0.10	0.29	0	1
West North Central	12,007	0.14	0.35	0	1
South Atlantic	12,007	0.19	0.39	0	1
East South Central	12,007	0.05	0.22	0	1
West South Central	12,007	0.10	0.30	0	1
Mountain North	12,007	0.04	0.19	0	1
Mountain South	12,007	0.03	0.17	0	1
Pacific	12,007	0.17	0.38	0	1
AIA Climate Zone					
≤ 2,000 CDD and ≥ 7,000 HDD	12,007	0.10	0.30	0	1
< 2,000 CDD and 5,500–7,000 HDD	12,007	0.20	0.40	0	1
< 2,000 CDD and 4,000–5,499 HDD	12,007	0.25	0.44	0	1
< 2,000 CDD and < 4,000 HDD	12,007	0.22	0.42	0	1
≥ 2,000 CDD and < 4,000 HDD	12,007	0.22	0.42	0	1
Metropolitan indicator					
Indicator for MSA	12,007	0.85	0.35	0	1
Indicator for urban	12,007	0.80	0.40	0	1

rion metrics indicate that the two-class model provides the best fit to the data. Based on these results, we proceed by using a two-class specification.

We also examine class selection from an energy efficiency perspective. As indicated in Section 3, the estimated efficiencies are likely to be biased within the entire sample because stochastic cost frontier estimates do not control for latent differences in household technologies. Based on the stochastic cost frontier estimates for the different classes, and the sample as a whole, we estimate

Table 3: Descriptive statistics—physical characteristics of housing

Variable	Obs	Mean	Std Dev	Min	Max
Characteristics of the home					
Total square feet of interior	12,007	2,174	1,451	100	16,122
Indicator for renter	12,007	0.31	0.46	0	1
Year residence was built	12,007	1,971	25	1,920	2,009
Number of refrigerators	12,007	1.27	0.51	0	7
Av. temperature of thermostat	12,007	70	4	40	90
Exterior wall type					
Brick	12,007	0.25	0.43	0	1
Wood	12,007	0.19	0.39	0	1
Siding (aluminum, vinyl, steel)	12,007	0.34	0.47	0	1
Stucco	12,007	0.15	0.36	0	1
Composite (shingle)	12,007	0.02	0.12	0	1
Stone	12,007	0.01	0.09	0	1
Concrete	12,007	0.05	0.21	0	1
General use of other fuels					
Natural gas	12,007	0.62	0.49	0	1
Heating oil	12,007	0.07	0.26	0	1
Liquid petroleum gas	12,007	0.43	0.50	0	1
Wood	12,007	0.12	0.32	0	1
Kerosene	12,007	0.01	0.12	0	1
Solar	12,007	0.01	0.11	0	1
Demand for energy conservation					
Well insulated home	12,007	0.36	0.48	0	1
Adequately insulated home	12,007	0.44	0.50	0	1
Poorly insulated home	12,007	0.19	0.40	0	1
Single-paned windows	12,007	0.42	0.49	0	1
Double-paned windows	12,007	0.56	0.50	0	1
Triple-paned windows	12,007	0.01	0.12	0	1
Home energy audit	12,007	0.05	0.22	0	1

efficiency indexes for each class. The descriptive statistics for the efficiency indexes are listed in Table 6. The overall sample mean value of energy efficiency suggests that households consume their energy resources relatively efficiently. That is, on average, households consume energy with 86% efficiency, relative to the theoretical cost-minimizing frontier. As the entire sample is stratified, the average efficiency level increases (incrementally) as predicted—this is arguably due to the fact that we are controlling for latent household-level energy technologies within each of the classes. The differences in estimated efficiencies between strata (and the sample mean) imply that we can reject the hypothesis that there are no differences in technical efficiencies between classes (what we label as "Hypothesis two" above). We further tested this hypothesis through a series of (*t*-tests) mean-comparison tests and reject the null hypothesis (no statistically significant difference between the mean values) in favor of the alternative. In other words, we find relatively strong evidence of statistically significant differences in mean efficiencies between each of the classes. The last estimated mean and standard deviation of the efficiency index, labeled in Table 6 as "Posterior two-class," is based on the class-level (posterior probability weighted) efficiency index offered in equation (20).

Casual observation of the differences in the average efficiency estimates may seem economically inconsequential; however, consider that the average household (within the sample) consumed approximately 90 million BTUs of energy over a one-year period. A difference in efficiency as small as one percent is approximately equal to 0.9 million BTUs of energy consumption. In 2014, the average residential lightbulb (converted from kilowatt-hours) consumed about 3,412 BTUs in

Table 4: Descriptive statistics—socio-economic characteristics

Variable	**Obs**	**Mean**	**Std Dev**	**Min**	**Max**
Socioeconomic data of occupants					
Female head of house	12,007	0.47	0.50	0	1
Retired	12,007	0.30	0.46	0	1
Social security income	12,007	0.08	0.27	0	1
Bachelor's degree	12,007	0.31	0.46	0	1
Number of persons residing	12,007	2.67	1.52	1	14
Poverty	12,007	0.14	0.35	0	1
Married	12,007	0.60	0.49	0	1
Racial makeup of occupants					
African American	12,007	0.13	0.33	0	1
White/Caucasian	12,007	0.79	0.41	0	1
Asian	12,007	0.04	0.19	0	1
Hispanic	12,007	0.14	0.35	0	1
Type of dwelling					
Mobile home	12,007	0.04	0.21	0	1
Apartment building	12,007	0.23	0.42	0	1
Single family home	12,007	0.72	0.45	0	1
Other characteristics					
Number of bedrooms	12,007	3	1	0	13
Use air conditioning	12,007	0.82	0.38	0	1
Space heating	12,007	0.96	0.19	0	1

Table 5: Prior class probability metrics—information criterion results

Model	Number of parms	BIC	AIC	CAIC	HQIC
2-class model	155	**–7484.09**	**–8644.83**	**–7327.09**	**–8255.48**
3-class model	244	–6359.50	–8185.63	–6112.50	–7573.07
4-class model	333	–5346.76	–7838.28	–5009.76	–7002.53
5-class model	422	–3883.90	–7040.82	–3456.90	–5981.87
6-class model	511	–3015.13	–6837.44	–2498.13	–5555.28

Notes: BIC denotes the (Schwartz) Bayesian Information Criterion; AIC denotes the Akaike Information Criterion; CAIC denotes denotes the Consistent Akaike Information Criterion; and, HQIC denotes the Hannan-Quinn Information Criterion. The model with the best fit to the data is determined by the lowest value (highlighted in bold) of each metric.

Table 6: Descriptive statistics of the efficiency indexes for the different class stratifications

Class	**Mean**	**Std Dev**
2	0.8680	0.0781
3	0.8705	0.0775
4	0.8715	0.0779
5	0.8725	0.0784
6	0.8715	0.0793
Full sample	0.8590	0.0786
Posterior two-class	0.8721	0.3305

Notes: The final efficiency index calculation is based on equation (20).

Table 7: Prior and posterior class probabilities (averages in decimal form) and other average class characteristics

Class	Number of households	Prior	Posterior	Total energy costs (US$)	Total energy consumption (thousand BTUs)	Total square feet
	6,004	0.5000	0.5226	1,854.20	74.81	2,703.09
	6,003	0.4999	0.4774	2,221.46	105.55	1,644.14
Entire sample	12,007	N/A	N/A	2,037.81	90.18	2,173.66

one hour (EIA, 2015a). Thus, a three-percent difference in estimate efficiency is roughly equivalent to 265 hours (0.9e+6/3412) of lightbulb usage throughout the year.

For the two-class specification, we display the prior and posterior class probabilities and other class characteristics, based on averages within the two separate classes. The prior and posterior probabilities are displayed in Table 7.

5. EMPIRICAL RESULTS

In this section we discuss the empirical results for the estimation procedure outlined in Section 3. We first estimate the stochastic frontier model using standard ordinary least squares (without restrictions) and estimate the various theoretically-consistent models including the latent-class stochastic frontier model, which we posit should provide the least biased estimates of technical inefficiency. In addition to providing the estimates from the espoused latent-class stochastic frontier model, we examine and measure the potential rebound effects. We will initially omit the conditioning factor estimates when displaying the regression results; however, in the final portion of this section we will more closely examine the conditioning factors—i.e., physical and socio-economic characteristics of the individual households—to ascertain why some units consume energy resources more (in)efficiently than others within the sample or class. However, it is important to first briefly explain the variable we used to estimate heteroskedasticity within the (in)efficiency term.

As shown in Figure 1, forty states have some type of energy building code regulations. However, fourteen of the states (and the District of Columbia) have more stringent regulations than the rest of the country. These fifteen jurisdictions have building code regulations that meet or exceed the 2012 International Energy Conservation Code standards. Notably, 2012 falls outside of the range of observation for our study, but we assume that if a jurisdiction required rigorous regulations in 2012, then it is likely the jurisdiction was equally as strict three years prior in 2009 (the temporal period of our observations). (These jurisdictions include California, Delaware, the District of Columbia, Florida, Illinois, Iowa, Maryland, Massachusetts, Michigan, Minnesota, Nevada, New Jersey, Rhode Island, Vermont, and Washington.) The descriptive statistics in Table 1 indicate that approximately forty percent of our sample falls within one of these jurisdictions.

5.1 Latent-class stochastic cost frontier model results

The parameter estimates for the stochastic cost frontier model are presented in Table 8. The table contains the frontier estimates for ordinary least squares (OLS), and the coefficients (and standard errors) in the third and fourth column are estimated by maximum likelihood. As a preliminary examination of the data, columns two and three of Table 8 offer the results for the entire sample (i.e., we assume the sample is composed of only one homogeneous class). The results of the

Table 8: Partial effects estimates for the cost frontier analysis (heteroskedastic error term)

Parameters	OLS	Full Sample	Class 1	Class 2
Cost frontier				
Total consumption	1.02***	1.00***	1.34***	0.01
	(0.07)	(0.05)	(0.07)	(0.11)
Total consumption sq'd	–0.06***	–0.06***	–0.09***	–0.04***
	(0.01)	(0.01)	(0.01)	(0.01)
Elec prices	1.64***	1.47***	1.74***	1.02***
	(0.14)	(0.09)	(0.12)	(0.15)
Nat gas prices	0.56***	0.65***	0.47***	0.74***
	(0.15)	(0.10)	(0.14)	(0.14)
Elec prices sq'd	–0.41***	–0.54***	–0.57***	–0.45***
	(0.05)	(0.02)	(0.03)	(0.03)
Nat gas prices sq'd	–0.15***	–0.19***	–0.13***	–0.23***
	(0.05)	(0.04)	(0.05)	(0.05)
Elec × nat gas prices	–0.04***	–0.03***	–0.03***	–0.02***
	(0.01)	(0.00)	(0.01)	(0.01)
Elec × consumption	–0.17***	–0.19***	–0.21***	–0.14***
	(0.01)	(0.01)	(0.01)	(0.01)
Nat gas × consumption	0.03***	0.04***	–0.03***	0.04***
	(0.00)	(0.00)	(0.00)	(0.00)
Constant	–21.27***	–21.73***		
	(4.71)	(4.24)		
Efficiency estimates				
σ^2_η				
Regulations	—	0.34***	0.54***	0.12***
	—	(0.04)	(0.06)	(0.05)
Constant	—	–3.20***	–3.68***	–3.10***
	—	(0.04)	(0.07)	(0.05)
σ^2_v				
Constant	—	–4.56***	–4.45***	–4.81***
	—	(0.04)	(0.05)	(0.06)
Average marginal effects of Regulations				
$E(\eta)$		0.03	0.04	0.01
Observations	12,007	12,007	6,004	6,003
R^2	0.92	—		

Notes: Standard errors in parentheses. The super scripted symbols denote the following *p*-values: '***' p < 0.01, '**' p < 0.05, and '*' p < 0.1. The additional control variables are included in the regression but the estimated coefficients of these variables are reported in Table 11 below. All variables are expressed in natural logarithms.

MLE (the stochastic cost frontier) specification suggests that state-level energy building codes play a role in household energy consumption, as the estimated parameter for the "Regulations" (within the variance of the inefficiency parameter, σ^2_η) is statistically significant at the one percent level. Moreover, the positive estimated sign on the term "Regulations" suggests that building codes, counter-intuitively, lead to an increase in total household-level energy expenditures. More specifically, the average marginal effects of "Regulations" implies that if a household is located in a state with rigorous energy building codes, then the house, on average, is likely to increase total energy expenditures by approximately three percent. (We explore this argument further below). The OLS specification, on the other hand, does not provide separate estimates for the noise term, v_i, and the inefficiency term, η_i, because OLS implicitly treats ε_i (the composite error term) as the stochastic error (and ignores any potential technical inefficiency). As such, the OLS specification should yield a normally distributed error term. However, the estimated skewness and kurtosis for the OLS model is 0.53 and 5.65 (not provided); and, a skewness/kurtosis hypothesis test for normality (i.e., the null hypothesis) was strongly rejected at a one percent significance level. The positive value of the skewness estimate implies that the error term is skewed to the right, which is consistent with the assumption of a cost frontier model (Kumbhakar et al., 2015).

For both specifications, the estimated cost frontiers are increasing in output (consumption), electricity, and natural gas input prices, as predicted by theory. The cost frontier implies concavity with respect to the input prices as the second derivatives of the frontier (with respect to input prices) are negative in both models. Additional analysis confirmed that both electricity and natural gas prices are positively monotonic with respect to the cost frontier. Further, an examination of the Hessian matrix reveals that the model satisfies the concavity condition for the input prices.

A reasonable postulate is that if a house is located within a jurisdiction that has stricter building code requirements, then it should, on average, lead to greater technical efficiency gains. Our analysis of the full sample reveals just the opposite—the estimated coefficient on the effect of regulations on technical efficiency is 0.34, which is statistically significant at the one percent level. This positive coefficient implies that regulations are leading to an increase in household energy costs. However, this estimated coefficient cannot be interpreted directly—the actual impact is based on the *marginal effect* of regulations on inefficiency (Kumbhakar et al., 2015). Based on this insight, the calculated marginal effect of regulations, on the mean of inefficiency, is approximately equal to 0.028. In other words, the estimate implies that stricter building codes, on average, lead to about a three-percent overuse of energy resources. The positive estimated effect may suggest a type of energy rebound effect, wherein the regulations require greater levels of energy efficiency but inadvertently lead to households consuming more energy because the effective cost of energy has decreased as a result of increased efficiency standards within the home. To get a more accurate depiction of the effect of regulations on technical inefficiency, we now turn to the class-level cost frontier analysis that should offer less biased estimates of technical inefficiency.

Columns four and five of Table 8 offer the parameter estimates for each of the latent class cost frontier models. That is, the data-driven method (outlined in Section 3.2) suggests that the underlying sample is best represented by the two-class model. The additional control variables, outside of the normal cost frontier variables displayed in the table, are omitted here for ease of exhibition, but the additional controls are offered in Table 11. The table contains two parts: the cost frontier estimates and the error terms. We do not impose the price homogeneity restrictions on the input prices, as we found relatively strong evidence of monotonicity and concavity before. (The discerning reader should not interpret this as a rejection of the homogeneity assumption. Rather, our maximum likelihood algorithm is complicated enough without having to impose any additional restrictions on the maximization problem, so we elected not to impose the homogeneity condition to improve computation). As predicted by theory, the estimated cost frontier elasticities are positively increasing in output (although, the estimate is only statistically significant for class one), electricity prices, and natural gas prices. The estimated technical efficiency term, σ_η^2, is highly statistically significant in both classes. The differences in estimated (in)efficiency across the classes lends support to our hypothesis of the underlying sample being characterized by unobserved heterogeneity.

The estimated cost frontiers also provide a measure of scale economies. Returns to scale (RTS) can be estimated, at the sample mean, as one minus the output cost elasticity: $RTS = 1 - \partial \ln C / \partial \ln q$, where q denotes the total amount of energy consumption. The RTS estimates for class two, although insignificant, are less than unity indicating the presence of increasing returns to scale. The RTS estimate for class one, on the hand, is less than zero indicating the presence of decreasing returns to scale. These results imply that, on average, there are other potential gains to be had (in energy savings) for households in class two. (We will explore these implications more thoroughly in the next subsection.) Neither of the class's output elasticity estimates exhibit constant returns to scale. Therefore, we can reject the constant returns to scale hypothesis (what we label above as "Hypothesis four") for each of the individual classes; however, we cannot reject hypothesis four for the entire sample (column four of Table 8).

We can also use the parameter estimates of Table 8 to test a Cobb-Douglas functional specification (or what we labelled as "Hypothesis three" above). The Cobb-Douglas restriction, to the translog cost function specification, is determined by the statistical significance of the coefficient on the cross-product of electricity and natural gas prices (or what is labelled in Tables 8 as "Elec × nat gas prices"). The Cobb-Douglas restriction can be strongly rejected as the cross-product estimate is highly statistically significant for both classes. These rejections imply that a Cobb-Douglas specification of the cost frontier model is too restrictive to represent the cost function for the underlying sample.

As observed above for the full sample, the average marginal effects of state-level residential energy building codes seems to induce higher energy costs for households in both classes. Such codes arguably induce a four percent and one percent increase in energy expenditures for households within class one and class two, respectively. However, before we accept the average marginal effects estimates it would be useful to further explore the potential for rebound effects using this particular methodology. In order to examine the rebound effects, we apply our results to a recently developed methodology offered by Orea et al. (2015).

5.2 A closer look at the rebound effects estimates

Orea et al. (2015) and Filippini and Hunt (2012) used a similar approach as the current study to analyze household energy efficiency. However, Orea et al. (2015) pointed out that a stochastic frontier model approach, to analyzing household energy consumption and efficiency, implicitly imposes a zero rebound effect for each household (at least when the efficiency parameter is assumed to follow a half-normal distribution). In order to deal with this potential limitation, Orea et al. (2015, p. 602) proposed augmenting the efficiency term (η) specification so that it is multiplicative with the term $(1-R)$, where R denotes a scalar parameter that measures the direction and magnitude of the rebound effect. Given this definition, the authors were able to develop two specific definitions of rebound effects based on the estimated sign of R. That is, if the effect of an improvement in energy use is attenuated by a "partial' rebound effect, then the estimated sign would be defined as: $0<R<1$. If, on the other hand, the effect of an energy efficiency improve is exacerbated then the estimated sign should be defined as $R<0$. Orea et al. (2015) define the former effect as a "partial rebound" and the latter as "super-conservative." In the context of our study, the idea behind a partial rebound is that residential energy building codes lead to technological changes, which reduce the effective price of energy resources, and therefore induce increases in energy consumption. (The authors also define the case of $R>1$, which they define as a "backfire effect;" however, the underlying assumptions within the stochastic cost frontier model preclude testing such an effect (Orea et al., 2015, p. 602)). Finally, if $R=0$, then it implies that there is no rebound effect whatsoever.

In addition to the basic definitions of rebound effects, Orea et al. (2015) outlined the limitations of estimating the augmented efficiency term (discussed in the previous paragraph). Basically, the researcher has to *a priori* restrict the parameters to estimate the augmented efficiency term. A discussion of the *a priori* restrictions is beyond the scope of the current paper; however, the reader is referred to Orea et al. (2015, p. 603) for a more detailed discussion of the restrictions. Neverthe-

Table 9: Energy efficiency index measures with and without the Orea et al. (2015) restrictions

	Mean	Std. dev.	Min.	Max.
Latent-class sfa (without restrictions)				
Class One	0.8790	0.6723	0.2921	0.9810
Class Two	0.8571	0.0891	0.2507	0.9910
Partial rebound effect restrictions				
Class One				
Adjusted	0.8683	0.0719	0.2885	0.9804
Non-adjusted	0.8659	0.0727	0.2823	0.9798
Class Two				
Adjusted	0.8524	0.0909	0.2461	0.0989
Non-adjusted	0.8490	0.0925	0.2376	0.9882

less, we produce the varying measures of rebound effects, based on the *a priori* restrictions, as a sensitivity analysis. As we found that state-level residential energy building codes arguably induce higher energy expenditures (on average), we limit our analysis to the unrestricted case and the partial rebound effect.

The rebound effect estimates appear in Table 9. As highlighted in the table, households within class one consumed energy more efficiently (on average) than did households within class two. The unrestricted efficiency index measures suggested that class one households consumed their energy services with approximately 88% efficiency (i.e., relative to the cost-minimizing frontier), whereas class two households consumed energy services with approximately 86% efficiency. However, the average efficiency (for both classes) dropped by approximately one percentage point when using the Orea et al. (2015) adjustments. Furthermore, the "Adjusted" partial rebound effect estimate suggests that households in class two consumed their energy services by approximately 15% (1–0.8524) more than the cost-minimizing frontier.

The results in Table 9 seem to offer further credence to the estimates found within Table 8. Specifically, we found that strict building codes seem to induce greater energy expenditures among households (holding all else equal). The adjustments in Table 9 also take these regulations into account, and as demonstrated by the efficiency scores for the restricted effects estimates, the average efficiency scores decreased when we control for state-level energy building codes.

5.3 Prior probability estimates and class-specific partial effects estimates

Table 10 offers the prior probability estimates, from (18), which are based on the multinomial logistic (MNL) regressions. The product of the prior class probabilities and likelihood function estimates were used to calculate the posterior probabilities of class membership, based on equation (19). The explanatory variables within MNL regression consist of: dummy variables for Census divisions; dummy variables for climate regions; dummy variables for AIA climate zones; and, heating and cooling degree days. These selected explanatory variables are arguably exogenous as they are determined outside of the control of the individual households. As indicated in Section 3, we selected exogenous variables to minimize the influence of the prior probability selection on the cost frontier model estimates (and ultimately yield unbiased estimates of technical (in)efficiency) (Greene, 2002; Brown et al., 2014).

Recall that we initially stratified (according to quantiles, which are labeled as "Quantile[s]" within the table) the entire sample according to the energy intensity (ratio of total energy consump-

Table 10: Prior probability estimates based on the multinomial logistic model

Multinomial logisitistic model	Quantile 1	Quantile 2
Census divisions		
Middle Atlantic		0.13
		(0.16)
East-North Central		0.24
		(0.15)
West-North Central		–0.09
		(0.16)
South Atlantic		–0.40***
		(0.15)
East-South Central		–0.52***
		(0.14)
West-South Central		–0.25
		(0.16)
Mountain-North		0.38***
		(0.15)
Mountain-South		–0.21
		(0.17)
Pacific		–0.15
		(0.16)
Climate regions		
Hot-dry/Mixed-dry		0.65***
		(0.18)
Hot-humid		0.47***
		(0.12)
Mixed-humid		0.28
		(0.19)
Marine		0.53***
		(0.17)
AIA Climate Zones		
< 2,000 CDD and 5,500–7,000 HDD		–0.52**
		(0.25)
< 2,000 CDD and 4,000–5,499 HDD		0.06
		(0.21)
< 2,000 CDD and < 4,000 HDD		0.30*
		(0.17)
≥ 2,000 CDD and < 4,000 HDD		0.17
		(0.14)
CDD		2e-4***
		(0.00)
HDD		2e-4***
		(0.00)
Constant		–1.76***
		(0.23)
Observations	12,007	12,007

Notes: Standard errors in parentheses. The superscript symbols denote the following *p*-values: '***' $p \leq 0.01$, '**' $p \leq 0.05$, and'*' $p \leq 0.1$. There are no multinomial logistic coefficient estimates for the first class because it is the baseline comparison group. "AIA" denotes the American Institute for Architecture.

tion to total square feet) of consumption within each household. Therefore, the first quantile (the baseline comparison group of analysis labeled as "Quantile 1") consisted of the least energy-intensive households, followed by quantile two. The coefficients within the second quantile represent the directional effect on prior class membership. (To be precise, we should arguably analyze the marginal effects instead of the coefficients to determine the probability of class membership. However, the sign of the marginal effects are the same as the estimated coefficients, and we are not particularly interested in the magnitude of the individual (marginal) effects, so we limit our analysis to the estimated coefficients only). A fairly large portion of the variables (Census division, climate region, climate zones, and degree days) are statistically significant indicating that the inclusion of the chosen variables conveys useful information in determining prior class probabilities, which in turn are used to determine the final class stratification. For example, the negative sign on the "South Atlantic" Census division variable (for quantile two) indicates that a household observation located within that particular region is less likely to belong to class two. Further, the sign on the climate region variable labelled as "Hot-dry/Mixed-dry" was also found to be positive for class two. The variable for cooling degree days (and heating degree days), which represents the demand for internal cooling (and internal heating), are positive and highly significant suggesting their inclusion provides useful information for class stratification.

The covariate estimates, used in addition to the theoretical cost frontier parameters, are displayed in Table 11. The parameter estimates for class one suggests that a ten percent increase in the square footage within a home would lead to a 0.2% decrease in energy costs. However, the class two estimates imply that a ten percent increase in footage would lead to a 0.1% increase in energy costs. Further, similar to the results in Levinson (2016), we found that newer homes, as identified by the coefficient estimates on "Year built" (Table 11) for classes one and two, arguably consume more energy. However, we also specify additional indicator variables for vintage of the home and estimates for the newest vintage (2004–2009) suggest that energy usage decreased marginally for households observed within class two.

The number of refrigerators (labeled as "Number of frigs" in Table 11) in a home is a fairly easy metric to determine how wasteful households are with their energy consumption, as some consumers have a tendency to purchase newer, energy-efficient refrigerators and keep older, less-efficient units for additional storage (Kim et al., 2006). We find consistent evidence that an increase in the number of refrigerators in the home leads to an increase in energy usage—the finding is also highly statistically significant across each of the classes. Consistent with our priors, the estimates imply that an increase in the number of bedrooms leads to an increase in energy consumption.

In addition to the two main energy inputs (electricity and natural gas) that we explored within the current study, the initial survey (RECS) provides data on the consumption of liquid petroleum gasoline, kerosene, heating oil, wood, and solar energy. (We did not include the additional energy resources within the main cost frontier parameters because the survey lacked a sufficient number of observations for the latter five sources. The indicator for heating oil formed the baseline comparison group for the other sources, so it was omitted from the regression results in Table 11). A close observer may also notice that we include a control for natural gas usage (labelled as "Use nat gas" in Table 11)—this was due to the fact that only approximately 60% of the survey respondents indicated that they used natural gas, as opposed to 100% of respondents who indicated that they used electricity. If a survey entry for natural gas consumption was zero, we did not know if the household truly consumed zero units of natural gas or if it was missing data coded as zero. Therefore, we include the additional indicator variable for natural gas consumption to control for potential data censoring issues (Burbidge et al., 1988). Nevertheless, the indicator suggests an inverse relationship between natural gas consumption and energy usage—these results are negative and highly statistically significant across each of the class-level estimates. A small portion of the survey respondents use solar energy (approximately one percent of the overall responses); however, the coefficient estimates are insignificant.

Table 11: Class-specific partial effects—covariates (heteroskedastic error)

Parameters	Class 1	Class 2	Parameters	Class 1	Class 2
Characterists of home			*Demographics*		
Total square ft	–0.02**	0.01*	Married	0.03***	0.04***
	(0.01)	(0.01)		(0.01)	(0.00)
Metropolitan	–0.00	–0.01**	Poverty	0.01	0.01
	(0.01)	(0.01)		0.01	0.01
Urban	–0.01**	–0.01**	Retired	–0.02***	–0.02***
	(0.01)	(0.01)		(0.01)	(0.01)
Renter	0.00	–0.01**	College educ	–0.00	–0.02***
	(0.01)	(0.01)		(0.00)	(0.01)
Year built	2.00**	3.05***	Caucasian	–0.01	–0.02***
	(0.88)	(0.70)		(0.01)	(0.01)
Number of frigs	0.04***	0.04***	Latino	0.01	0.02***
	(0.00)	(0.01)		(0.01)	(0.01)
Number of beds	0.07***	0.09***	African American	0.00	0.01
	(0.01)	(0.01)		(0.01)	(0.01)
Single family	–0.01	0.03***	Asian	–0.00	–0.03***
	(0.01)	(0.01)		(0.01)	(0.01)
			No. of occupants	0.00	–0.00
Vintage of home				(0.00)	(0.00)
1950s	–0.01	–0.02*	Soc sec income	0.03***	0.02***
	(0.01)	(0.02)		(0.00)	(0.01)
1960s	0.02	–0.03*	HH age	–0.00	–0.01***
	(0.02)	(0.01)		(0.00)	(0.00)
1970s	0.01	–0.02	HH male	–0.00	–0.01***
	(0.02)	(0.02)		(0.00)	(0.00)
1980s	0.01	–0.03			
	(0.03)	(0.02)	*Income*		
1990s	–0.00	–0.03	\$20–40K	0.01	0.00
	(0.03)	(0.02)		(0.01)	(0.01)
2000–2004	–0.00	–0.04	\$40–60K	0.02*	0.01*
	(0.03)	(0.03)		(0.01)	(0.01)
2004–2009	–0.03	–0.07**	\$60–80K	0.03**	0.01
	(0.04)	(0.03)		(0.01)	(0.01)
			\$80–100K	0.02*	0.03***
Energy resource use				(0.01)	(0.01)
Use liq pet gas	0.00	–0.02***	≥ \$100K	0.05***	0.04***
	(0.00)	(0.01)		(0.01)	(0.01)
Use nat gas	–2.29***	–2.80***			
	(0.17)	(0.17)	*Conservation indicators*		
Use kerosene	–0.01	–0.03**	Use air cond	0.07***	0.07***
	(0.02)	(0.02)		(0.01)	(0.01)
Use wood	–0.02***	0.01	Thermostat	0.23***	0.18***
	(0.01)	(0.01)		(0.03)	(0.03)
Use solar	0.00	0.03	Well insulated	–0.02	–0.03
	(0.02)	(0.02)		(0.02)	(0.02)
			Adeq insulated	–0.02	–0.03
				(0.02)	(0.02)
			Poorly insulated	–0.02	–0.02
				(0.02)	(0.02)
			1-pane glass	0.00	0.01
				(0.03)	(0.03)
			2-pane glass	0.02	0.01
				(0.03)	(0.03)
			3-pane glass	0.01	0.02
				(0.03)	(0.03)
			Energy audit	0.01	0.01
				(0.00)	(0.01)
Observations	6,004	6,003	Observations	6,004	6,003

Notes: Standard errors in parentheses. The super scripted symbols denote the following *p*-values: '***' $p < 0.01$, '**' $p < 0.05$, and '*' $p < 0.1$. All variables are expressed in natural logarithms.

Columns four, five, and six of Table 11 highlight additional demographic and conservation-related covariates. The coefficient estimates indicate that if a survey respondent is married, he or she is more likely to consume additional energy within the home—the marginal effect is small, but highly statistically significant across both of the classes. The coefficient estimates suggest that if a respondent were to switch from unmarried to married, then we could expect a three-to-four percent (approximately US$50–$70) annual increase in the geometric mean of energy costs (approximately US$ 1755 for the entire sample) within the home. If a respondent was retired at the time of the survey, then we would expect a relatively small two percent decrease (in the geometric mean) in energy consumption. Moreover, if a respondent is college educated ("College educ"), then it only leads to small marginal decrease in energy consumption, which is only significant for class two. The respondent's race (or ethnic heritage) does not seem to provide a clear pattern of an overall increase or decrease in energy consumption; although, householders of Latino or Asian descent consume marginally less energy relative to the other ethnicities represented in the survey. Consistent with expectations, the estimates suggest that the older the age ("HH age") of the householder (the respondent to the survey), the less energy is consumed within the home—this possibly could be due to older respondents living on fixed incomes after retirement age. The expected sign and magnitude of the estimated coefficients for "HH age" are similar to the estimate coefficients for the "Retired" indicator.

Among the income indicators in Table 11, only the group of respondents with the highest incomes (equal to or exceeding US$ 100K per year) seem to offer a relatively clear pattern of energy consumption. That is, if a respondent was at the highest income bracket at the time of the survey, we would expect to see an approximate five percent annual increase in the geometric mean (approximately $88) of energy costs. This may imply that wealthier respondents are less sensitive to energy input prices, and therefore consume more energy (relative to the cost-minimizing frontier). It could also be argued that wealthier respondents own larger living spaces, which require more energy services. An auxiliary regression of total energy costs (in natural logs) on an interaction variable of $100K-per-year respondents and the total number of household bedrooms (in natural logs) indicates that a ten percent increase (in the interaction term) leads to an approximate 4% increase in average energy costs. Our results suggest that wealthier respondents demand larger homes and by extension consume more energy services.

Beyond demographic variables, we explore indicators for conservation behavior among households—the estimation results are presented at the bottom of Table 11. Consistent with expectations, if a household uses air conditioning (labelled as "Use air cond"), then overall household-level energy costs increase. The estimates imply that the use of air conditioning raises the geometric mean of average energy costs by seven percent per year. The intuition with a household's average thermostat setting (labelled as "Thermostat') is similar to that of air conditioning—that is, all else equal, a ten percent increase in the average thermostat setting (to control internal heating and cooling) leads to an approximate two percent increase in average annual energy costs. The thermostat variable should be interpreted with skepticism as this is a self-reported value as opposed to an average annual reading from an auditor. Surprisingly, a household's level of insulation and type of external windows does not appear to have a statistically significant affect on average annual energy costs. Again though, the insulation indicators should be viewed with skepticism as the responses are self-reported. Finally, we include an indicator if a home had received an energy audit (labelled as "Energy audit") in the recent past. Inconsistent with our expectations, the variable for energy audit(s) is statistically insignificant.

6. CONCLUSIONS AND POLICY IMPLICATIONS

The empirical results within the study yield many potential policy implications for U.S. residential energy building codes. One of the main findings is that recently promulgated energy building code regulations appear to have led to an increase, instead of a decrease, in average annual energy consumption. The current study—based on economic, not engineering analysis—sheds light on the fact that potential savings may change due to behavior responses. These findings are fairly robust to different demographic characteristics of occupants, physical characteristics of the home, geographic regions, climate data, and unobserved heterogeneity within the underlying technologies (for energy services) available in the home. The increase in average annual consumption could suggest a rebound effect, wherein stricter energy building code requirements lead to greater energy efficiency within the household and thereby reduce the effective costs of energy resources. The reduction in the effective costs could lead to an unintended consequence in which households end up consuming more energy resources as a result of the inducement to retrofit the home. Our findings are not entirely bleak, as we do find some evidence where energy code regulations serve their purported intent by reducing overall average energy consumption for some of the observations. From a regulatory standpoint, it is important for policymakers to further investigate why some codes lead to an increase in energy consumption whereas others lead to a decrease.

There are some potential shortcomings within the current study stemming from the limitation of the underlying data. The RECS survey did not explicitly collect data on the average costs of the energy inputs (such as the average rate of electricity service per KWh); therefore, our study is limited by inferring these costs. Additionally, our analysis disregarded allocative inefficiency within household energy consumption. Future research may consider estimating both technical and allocative inefficiency of consumption. Further, future studies within this literature may consider offering a system of equations that includes additional theoretically-consistent assumptions in regards to cost minimization. For example, in addition to the cost frontier model, researchers can also include the cost-share equations for each of the inputs. As we found fairly consistent evidence that cost frontier analysis satisfied the assumptions of monotonicity and concavity, we did not pursue the systems equation approach. Lastly, Brown et al. (2014) recently criticized the use of multinomial logistic models in latent class analysis. Part of the reason is due to the undesirable assumption of the Independence from Irrelevant Alternatives—wherein it is assumed that the probability of class membership, relative to any other, is independent of any additions to, or deletions from, the choice set (Train, 1993). Brown et al. (2014) propose using an ordered probit instead of the multinomial logistic model. In cases like the current study, where the ordering matters for potential class membership, such an approach may prove to be fruitful for future research. As the ordered probit approach would have added only more complexity to the current study, we leave it for future research.

As governmental agencies and research institutions continue to gather rich microdata sets, economic theory (including stochastic frontier analysis) can provide several unique insights into household-level energy consumption and the efficiency of consumption, including the behavioral responses of households to increasingly stringent building codes. This differs from energy efficiency estimates based on engineering analysis, which often assumes away the behavioral aspects of household energy consumption. Thus, economic theory can aid in policy development for future energy efficiency standards or goals.

ACKNOWLEDGMENTS

This study greatly benefited from discussions with Carl Pasurka, among others. We would also like to thank two blind reviewers for offering insightful comments and suggestions.

REFERENCES

Aigner, D., C. A. K. Lovell, and P. Schmidt (1977). "Formulation and estimation of stochastic frontier production function models." *Journal of Econometrics* 6: 21–37. https://doi.org/10.1016/0304-4076(77)90052-5.

Aldenderfer, M. and R. Blashfield (1984). *"Cluster Analysis."* Washington, D.C: Sage Publications, Inc. https://doi.org/10.4135/9781412983648.

Allcott, H. (2013). "The welfare effects of misperceived product costs: Data and calibrations from the automobile market." *American Economic Journal: Economic Policy* 5(3): 30–66. https://doi.org/10.1257/pol.5.3.30.

Alvarez, A., C. Amsler, L. Orea, and P. Schmidt (2006). "Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics." *Journal of Productivity Analysis* 25(3): 201–212. https://doi.org/10.1007/s11123-006-7639-3.

Alvarez, A. and J. del Corral (2010). "Identifying different technologies using a latent class model: extensive versus intensive dairy farms." *European Review of Agricultural Economics* 37(2): 231–250. https://doi.org/10.1093/erae/jbq015.

Anderson, S. T. and R. G. Newell (2004). "Information programs for technology adoption: The case of energy-efficiency audits." *Resource and Energy Economics* 26(1): 27–50. https://doi.org/10.1016/j.reseneeco.2003.07.001.

Aroonruengsawat, A., M. Auffhammer, and H. Sanstad (2012). "The impact of state level building codes on residential electricity consumption." *The Energy Journal* 33(1): 31–52. https://doi.org/10.5547/ISSN0195-6574-EJ-Vol33-No1-2.

Baxter, L. W., S. L. Feldman, A. P. Schinnar, and R. M. Wirtshafter (1986). "An efficiency analysis of household energy use." *Energy Economics* 8(2): 62–73. https://doi.org/10.1016/0140-9883(86)90031-9.

Beard, T. R., S. B. Caudill, and D. M. Gropper (1991). "Finite mixture estimation of multiproduct cost functions." *The Review of Economics and Statistics* 73(4): 654–664. https://doi.org/10.2307/2109404.

Becker, G. S. (1981). *"Treatise on the family."* Cambridge, MA.: Harvard University Press.

Bozdogan, H. (1987). "Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions." *Psychometrika* 52(3): 345–370. https://doi.org/10.1007/BF02294361.

Building Code Assistance Project (2016). "State Residential Energy Code Status." Accessed online at http://bcap-energy.org/code-status/residential/, March 2016.

Brent, D. A., J. H. Cook, and S. Olsen (2013). "Hetergeneous responses to social norms for water conservation." Selected Paper for presentation at the Agricultural & Applied Economics Associations 2013 Annual Meeting. Accessed online at http://ageconsearch.umn.edu/bitstream/149697/2/Brent_WaterSmart_AAEA.pdf, March 2016.

Brown, S., W. H Greene, and M. N. Harris (2014). "A new formulation for latent class models." IZA Discussion Paper Series, Number 8283. Accessed online at http://ftp.iza.org/dp8283.pdf, November 2015.

Burbidge, J. B., L. Magee, and A. L. Robb (1988). "Alternative transformations to handle extreme values of the dependent variable." *Journal of the American Statistical Association* 83(401): 123–127. https://doi.org/10.1080/01621459.1988.10478575.

California Energy Commission (2013). "California Energy Commission: 2012 Accomplishments." Accessed online at http://www.energy.ca.gov/releases/2013_releases/2012_Accomplishments.pdf, January 2015.

Caudill, S. B. and J. M. Ford (1993). "Biases in frontier estimation due to heteroscedasticity." *Economics Letters* 41: 17–20. https://doi.org/10.1016/0165-1765(93)90104-K.

Caudill, S. B., J. M. Ford, and D. M. Gropper (1995). "Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity." *Journal of Business & Economic Statistics* 13(1): 105–111. Statistics 13(1): 105–111..

Chong, H. (2012). "Building vintage and electricity use: Old homes use less electricity in hot weather." *European Economic Review* 56(5): 906–930. https://doi.org/10.1016/j.euroecorev.2012.02.009.

Christensen, L. R. and W. H. Greene (1976). "Economies of scale in U.S. electric power generation." *The Journal of Political Economy* 84(4):655–676. https://doi.org/10.1086/260470.

Coelli, T. J., S. Singh, and E. Fleming (2003). "An input distance function approach to the measurement of technical and allocative efficiency: with application to Indian dairy processing plants." University of Queensland Working Paper. Accessed online at https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=esam2003 paper_id=105, April 2015.

Davis, L. W., A. Fuchs, and P. Gertler (2014). "Cash for coolers: Evaluating a large-scale appliance replacement program in Mexico." *American Economic Journal: Economic Policy* 6(4): 207–238. https://doi.org/10.1257/pol.6.4.207.

Farrell, M.J. (1957). “The measurement of productive efficiency.” *Journal of the Royal Statistical Society* 120(3): 253–290. https://doi.org/10.2307/2343100.

Filippini, M. and L. C. Hunt (2012). “US residential energy demand and energy efficiency: A stochastic demand frontier approach.” *Energy Economics* 34(5): 1484–1491. https://doi.org/10.1016/j.eneco.2012.06.013.

Fowlie, M., M. Greenstone, and C. Wolfram (2015). “Do energy efficiency investments deliver? Evidence from the Weatherization Assistance Program.” E2e Working Paper 020. Accessed online at http://e2e.haas.berkeley.edu/pdf/workingpapers/WP020.pdf#page=1, July 2015. https://doi.org/10.3386/w21331.

Gelman, A. and G. Imbens (2013). “Why ask why? Forward causal inference and reverse causal questions.” The National Bureau of Economic Research, NBER Working Paper No. 19614. Accessed online at http://www.nber.org/papers/w19614, February 24, 2016. https://doi.org/10.3386/w19614.

Gerarden, T., R. G. Newell, and R. N. Stavins (2015). “Deconstructing the energy-efficiency gap: Conceptual frameworks and evidence.” *American Economic Review* 105(5): 183–186. https://doi.org/10.1257/aer.p20151012.

Gillingham, K., M. Harding, and D. Rapson (2012). “Split incentives in residential energy consumption.” *The Energy Journal* 33(2): 37–62. https://doi.org/10.5547/01956574.33.2.3.

Greene, D. L., D. H. Evans, and J. Hiestand (2013). “Survey evidence on the willingness of U.S. consumers to pay for automotive fuel economy.” *Energy Policy* 61: 1539–1550. https://doi.org/10.1016/j.enpol.2013.05.050.

Greene, W. H. (2002). “Alternative panel data estimators for stochastic frontier models.” Working Paper, Department of Economics, Stern School of Business, NYU. Accessed online at www.stern.nyu.edu/ wgreene/panelfrontiers.doc, February 2016.

Greene, W. H. (2003). *“Econometric Analysis.”* Upper Saddle River, NJ: Prentice Hall, 5th edition.

Greene, W. H. (2005). “Reconsidering heterogeneity in panel data estimators of the stochastic frontier model.” *Journal of Econometrics* 126: 269–303. https://doi.org/10.1016/j.jeconom.2004.05.003.

Greenstone, M. and T. Gayer (2008). “Quasi-experimental and experimental approaches to environmental economics.” *Journal of Environmental Economics and Management* 57: 21–44. https://doi.org/10.1016/j.jeem.2008.02.004.

Griliches, Z. (1957). “Specification bias in estimates of production function.” *Journal of Farm Economics* 39: 8–20. https://doi.org/10.2307/1233881.

Hannan, E. and B. Quinn (1979). “The determination of the order of an autoregression.” *Journal of the Royal Statistical Society* B(41): 190–195.

Hinrichs, R. A. and M. Kleinbach (2006). *“Energy: Its use the environment.”* New York City: Brooks/Cole, Cengage Learning.

Ironmonger, D. (2000). “Household production and the household economy.”, The University of Melbourne, Department of Economics Research Paper. Accessed online at http://fbe.unimelb.edu.au/__data/assets/pdf_file/0009/805995/759.pdf, June 2015.

Ito, K., T. Ida, and M. Tanaka (2015). “The persistence of moral suasion and economic incentives: Field experimental evidence from energy demand.” NBER Working Paper No. 20910. Accessed online at http://www.nber.org/papers/w20910, March 2016. https://doi.org/10.3386/w20910.

Jacobsen, G. D. and M. J. Kotchen (2013). “Are building codes effective at saving energy? Evidence from residential billing data in Florida.” *The Review of Economics and Statistics* 95(1): 34–49. https://doi.org/10.1162/REST_a_00243.

Jaffe, A. B. and R. N. Stavins (1994). “Energy-efficient investments and public policy.” *Energy Policy* 22(1): 805–810.

Joskow, P. L. and D. B. Marron (1992). “What does a negawatt really cost? Evidence from utility conservation programs.” *The Energy Journal* 13(4): 41–74. https://doi.org/10.5547/ISSN0195-6574-EJ-Vol13-No4-3.

Kim, H. C., G. A. Keoleian, and Y. A. Horie (2006). “Optimal household refrigerator replacement policy for life cycle energy, greenhouse gas emissions, and cost.” *Energy Policy* 34(15): 2310–2323. https://doi.org/10.1016/j.enpol.2005.04.004.

Kumbhakar, S. C., H.-J. Wang, and A. P. Horcastle (2015). *“A Practioner’s Guide to Stochastic Frontier Analysis Using Stata.”* Cambridge, MA: Cambridge University Press. https://doi.org/10.1017/CBO9781139342070.

Kumbhakar, S. C., H.-J. Wang, and A. P. Horcastle (2016). “A Practioner’s Guide to Stochastic Frontier Analysis Using Stata.” Accessed online at https://sites.google.com/site/sfbook2014/, January 2016.

Laustsen, J. (2008). “Energy Efficiency Requirements in Building Codes, Energy Efficiency Policies for New Buildings.” International Energy Agency, IEA Information Paper. Accessed online at https://www.iea.org/publications/freepublications/publication/Building_Codes.pdf, March 2016.

Levinson, A. (2016). “How much energy do building energy codes really save? Evidence from California houses.” *American Economic Review* 106(1): 2867–2894. https://doi.org/10.1257/aer.20150102.

Lovell, C. A. K. (1993). “The Measurement of Productive Efficiency,” In A.Q. Fried, C.A.K. Lovell, and S.S. Schmidt, eds., *“Production frontiers and productive efficiency.”* New York City: Oxford University Press. pp. 3–67.

McKinsey and Company (2007). "Reducing U.S. greenhouse gas emissions: How much at what cost?" Accessed online at http://www.mckinsey.com/client_service/sustainability/latest_thinking/reducing_us_greenhouse_gas_emissions, January 2015.

Meeusen, W. and J. van Den Broeck (1977). "Efficiency estimation from Cobb-Douglas production functions with composed error." *International Economic Review* 18(2): 435–444. https://doi.org/10.2307/2525757.

Metcalf, G. E. and K. A. Hassett (1999). "Measuring The Energy Savings From Home Improvement Investments: Evidence From Monthly Billing Data." *The Review of Economics and Statistics* 81(3): 516–528. https://doi.org/10.1162/003465399558274.

Newell, R. G. and J. Siikamäki (2014). "Nudging energy efficiency behavior: The role of information labels." *Journal of the Association of Environmental and Resource Economists* 1(4): 555–598. https://doi.org/10.1086/679281.

Newman, D. K. and D. Day (1975). *"The American Energy Consumer."* Cambridge, MA: Ballinger.

Orea, L. and S. C. Kumbhakar (2004). "Efficiency measurement using a latent class stochastic frontier model." *Empirical Economics* 29(1): 169–183. https://doi.org/10.1007/s00181-003-0184-2.

Orea, L., M. Llorca, and M. Filippini (2015). "A new approach to measuring the rebound effect associated to energy efficiency improvements: An application to the US residential energy demand." *Energy Economics* 49: 599–609. https://doi.org/10.1016/j.eneco.2015.03.016.

Pollak, R. A. (2003). "Gary Becker's contributions to family and household economics." *Review of Economics of the Household* 1: 111–141. https://doi.org/10.1023/A:1021803615737.

Reifschneider, D. and R. Stevenson (1991). "Systematic departures from the frontier: A framework for the analysis of firm inefficiency." *International Economic Review* 32(3): 715–723. https://doi.org/10.2307/2527115.

Sallee, J. M. (2013). "Rational inattention and energy efficiency." National Bureau of Economic Research, Working Paper 19545. https://doi.org/10.3386/w19545.

Scott, A. (1980). "The economics of house heating." *Energy Economics* 2: 130–141. https://doi.org/10.1016/0140-9883(80)90024-9.

Socolow, R. H. (1978). "The Twin Rivers Program on energy conservation in housing: Highlights and conclusions." In R.H. Socolow, ed., *"Saving Energy in the Home."* New York City: Ballinger. https://doi.org/10.1016/0378-7788(78)90003-8.

Train, K. (1993). "Qualitative choice analysis: Theory, econometrics, and an application to automobile demand." Cambridge, MA: The MIT Press.

U.S. Department of Energy (1981). "Residential energy consumption survey 1979–1980 consumption and expenditures, Part I: National data." Energy Information Administration, DOE/EIA 0262/1, Washington, D.C.

U.S. Department of Energy (1981). "Residential energy consumption survey. 1979–1980 consumption and expenditures, Part II: Regional data." Energy Information Administration, DOE/EIA 0262/2, Washington, D.C.

U.S. Department of Energy (2014). "State Energy Program Impacts: Energy and Cost Savings from the State Energy Program." Accessed online at http://energy.gov/eere/wipo/state-energyprogram-impacts, April 2015.

U.S. Energy Information Administration (2012). "Residential Energy Consumption Survey (RECS): The impact of increasing home size on energy demand." Accessed online at https://www.eia.gov/consumption/residential/reports/2009/square-footage.cfm, March 2015.

U.S. Energy Information Administration (2015). "Drivers of U.S. Household Energy Consumption, 1980–2009." Accessed online at http://www.eia.gov/analysis/studies/buildings/households/, February 2016.

U.S. Energy Information Administration (2015). "Energy units and calculators explained." Accessed online at http://www.eia.gov/energyexplained/?page=about_energy_units, February 2016.

U.S. Environmental Protection Agency (2015). "Energy and environment guide to action: State policies and best practices for advancing energy efficiency, renewable energy, and combined heat and power." Accessed online at https://www.epa.gov/statelocalclimate/energy-and-environment-guide-action-full-report, April 2015.

Varian, H. R. (1992). *"Microeconomic analysis."* W. W. Norton & Company, New York City, 3rd edition.

Wang, H.-J. and P. Schmidt (2002). "One-Step and two-step estimation of the effects of exogenous variables on technical efficiency levels." *Journal Productivity Analysis* 18: 129–144. https://doi.org/10.1023/A:1016565719882.