Data-Driven Benchmarking of Building Energy Efficiency Utilizing Statistical Frontier Models

Amir Kavousian1 and Ram Rajagopal2

Abstract: Frontier methods quantify the energy efficiency of buildings by forming an efficient frontier (best-practice technology) and by comparing all buildings against that frontier. Because energy consumption fluctuates over time, the efficiency scores are stochastic random variables. Existing applications of frontier methods in energy efficiency either treat efficiency scores as deterministic values or estimate their uncertainty by resampling from one set of measurements. Availability of smart meter data (repeated measurements of energy consumption of buildings) enables using actual data to estimate the uncertainty in efficiency scores. Additionally, existing applications assume a linear form for an efficient frontier; i.e., they assume that the best-practice technology scales up and down proportionally with building characteristics. However, previous research shows that buildings are nonlinear systems. This paper proposes a statistical method called stochastic energy efficiency frontier (SEEF) to estimate a bias-corrected efficiency score and its confidence intervals from measured data. The paper proposes an algorithm to specify the functional form of the frontier, identify the probability distribution of the efficiency score of each building using measured data, and rank buildings based on their energy efficiency. To illustrate the power of SEEF, this paper presents the results from applying SEEF on a smart meter data set of 307 residential buildings in the United States. SEEF efficiency scores are used to rank individual buildings based on energy efficiency, to compare subpopulations of buildings, and to identify irregular behavior of buildings across different time-of-use periods. SEEF is an improvement to the energy-intensity method (comparing k · Wh/sq.ft:): whereas SEEF identifies efficient buildings across the entire spectrum of building sizes, the energy-intensity method showed bias toward smaller buildings. The results of this research are expected to assist researchers and practitioners compare and rank (i.e., benchmark) buildings more robustly and over a wider range of building types and sizes. Eventually, doing so is expected to result in improved resource allocation in energy-efficiency programs.

DOI: 10.1061/(ASCE)CP.1943-5487.0000327. © 2014 American Society of Civil Engineers.

Author keywords: Energy efficiency; Stochastic ranking; Data-driven benchmarking; Data envelopment analysis; Energy-efficient frontier.

Introduction

Quantifying building energy efficiency is essential to developing and monitoring energy-efficiency scores programs (Ahn et al. 2010). Two categories of methods for quantifying building energy efficiency are engineering and empirical methods (Lutzenhiser et al. 2010).

Engineering methods simulate building performance by creating thermodynamic models of building operations (Clevenger et al. 2012). These models utilize buildings’ design information and industry averages of fixtures to predict energy usage (Xydis et al. 2008). Depending on the level of detail of the model, availability of reliable data, quality of assumptions, and accuracy of the simulation engine, engineering models are very accurate in identifying sources of inefficiency in a building (Griffith and Crawley 2006; Pisello et al. 2012; Waltner and Goldstein 2012; D. Goldstein, personal communication, 2012). However, such detailed models require significant time and skills to build and maintain and, hence, are cost prohibitive in most instances (Patterson 1996; Yezioro et al. 2008; Maile 2010).

Empirical methods utilize measured energy consumption data to create models using variables such as climate, building characteristics, energy fixtures (e.g., appliances and equipment), and occupancy. Most empirical methods (such as the Energy Star method) depend on regression models and, hence, are limited by the shortcomings of ordinary regression (e.g., Atkinson 1982; Riani and Atkinson 2000; Savage 2009). Regression methods identify the average trend over the entire population, and then compare each building with that overall trend. Thus, a few badly performing buildings that skew the regression line toward less-efficient buildings may make a building with modest performance appear highly efficient. In contrast, nonparametric empirical benchmarking models such as data envelopment analysis (DEA—Farrell 1957) compare a building with the best-performing building in its own class. Such a comparison of a building with the best-practice technology in its own class (or of similar functionality) enables estimating the total potential for energy efficiency. The goal is to estimate the amount of a reduction in energy consumption is possible without compromising the functionality of the building.

In DEA model, each building’s class is quantified using a measure of its output capacity, called level of service (LOS) in this paper. LOS is an indicator of the utility (service, output) that a building provides to its occupants. LOS is defined in many ways, such as floor area, number of occupants, a weighted combination of energy fixtures, or the physical output of a building (e.g., number of products produced in a manufacturing plant). Once the LOS is...
defined, the efficient frontier in the DEA model specifies the lowest amount of energy for a given LOS (Fig. 1).

Frontier methods such as DEA quantify energy efficiency by examining how much energy a building consumes in relation to the service that it provides. Frontier methods are divided into two categories: parametric and nonparametric. Parametric frontier methods, also called stochastic frontier analysis (SFA), assume the presentation of the efficient frontier using a mathematical formula plus a stochastic error term called a random component (Aigner et al. 1977). Nonparametric frontier methods such as DEA make no assumptions about the mathematical formula of the frontier and, instead, form the frontier using measured data. The stochastic energy efficiency frontier (SEEF) model is based on the nonparametric assumption, making it an extension of the DEA model.

The remainder of this paper first explains the DEA model and its applications in the field of energy-efficiency scores, and then introduces the SEEF model and explains its innovations over existing methods. SEEF is applied to a smart meter data set, and the insights obtained through its use are noted to describe the power of the model.

**Related Work**

Previous studies applied DEA to estimate the energy efficiency of buildings. For example, Carnes et al. (1998) applied DEA to examine the energy efficiency of Texas State universities. They used electricity, fuel, and thermal energy as energy inputs, and used course credits taught and research expenditures as the outputs of energy consumption for these universities. They discussed several variations of DEA, including the use of multiple outputs and multiple inputs (e.g., nonenergy inputs such as capital investment). Onut and Soner (2006) used DEA to analyze the energy efficiency of five-star hotels in Antalya, Turkey and made recommendations for increasing energy efficiency. Grosche (2009) applied DEA to the U.S. Residential Energy Consumption Survey (RECS) data for increasing energy efficiency. Grosche (2009) applied DEA to the U.S. Residential Energy Consumption Survey (RECS) data

**Table 1**: Different Returns to Scale Technologies and Their Respective LPP Constraints and Frontier Shapes

<table>
<thead>
<tr>
<th>Constraint for $\sum_{i=1}^{k} z_i$ term</th>
<th>Constant returns to scale</th>
<th>Decreasing returns to scale</th>
<th>Variable returns to scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontier shape (in Fig. 2)</td>
<td>None</td>
<td>Points $O, C, D, E$ and the ray $EH_E$ and the $x$-axis</td>
<td>$A', A, B, C, D, E$ plus the ray $EH_E$ and the $x$-axis starting at $A'$</td>
</tr>
</tbody>
</table>

Fig. 1. Schematic of a (a) regression model; (b) DEA model

Fig. 2. Different functional forms for the efficient frontier

DEA Model

Consider $k$ buildings, each consuming a vector of $n$ energy sources $x_j \in \mathbb{R}_+^n$ to provide a vector of $m$ services $s_j \in \mathbb{R}_+^m$. The matrix of observed services is denoted by $S_{(k,m)}$ and the matrix of observed inputs is denoted by $X_{(k,n)}$.

A linear programming problem (LPP) such as LPP 1 calculates the efficiency values of buildings given input matrix $X$ and service matrix $S$ (Farrell 1957). For illustration purposes, assume the LPP corresponding to linear frontier is as follows (ray OH in Fig. 2):

$$\theta_j(s, x) = \min \lambda$$

subject to

$$zS \geq s \quad [\text{LPP1} = \text{LPP}(S, X)]$$

$$zX \leq \lambda x$$

$$z \in \mathbb{R}_+^k$$

The efficiency of observation $P$ is given by the solution to the LPP (1); i.e., $\theta_j(s, x)$, which is also equal to the ratio $S_P Q / S_P P$. In other words, $\lambda$ is the ratio that enables a reduction in the energy consumption of a building without reducing its LOS. Minimizing $\lambda$ enables determining the maximum achievable energy reduction without reducing the LOS.

The value of $z$ is the intensity vector and serves to construct the smallest cone that covers the data in the sample. Adding constraints to the values of $z$ changes the shape of the frontier and the efficiency values. These constraints depend on the assumptions about the relationship between energy consumption and LOS, and are called returns to scale technologies. The term returns to scale is used frequently in economics to describe the relationship between changes in inputs versus outputs. For instance, decreasing returns to scale (DRS) implies that the output (in our case LOS) grows at a slower rate than that of input (in our case energy consumption). A constant returns to scale (CRS) implies that the LOS and energy consumption increase or decrease at the same rate. Finally, a variable returns to scale (VRS) indicates that the rate of change of LOS is higher or lower than that of energy consumption depending on the amount of energy consumption. Table 1 lists different assumptions for returns to scale and their corresponding LPP constraints.
from 1997 and 2001 and compared the energy efficiency of U.S. residential building stock during that period. Grosche (2009) concluded that the energy efficiency in U.S. households improved from 1997 and 2001 but that most of that improvement came from households using natural gas and fuel oil for their heating purposes.

One limitation of studies that applied DEA to energy efficiency is that they all used a linear parametric frontier without a data-driven validation of the frontier shape. This paper later shows how the shape of the frontier is validated using the data.

Another limitation to existing studies of DEA for energy efficiency is the assumption of deterministic energy-efficiency scores. Before the widespread use of high-frequency meters, studies that analyzed energy consumption used a limited data set of energy consumption, usually available for only a short period. Given a lack of repeated measurements, these studies assumed a deterministic consumption level for each building. However, energy consumption has a significant stochastic component, affected by factors such as weather and occupancy patterns (Azar and Menassa 2012; Berges et al. 2011; Chen et al. 2012). Because DEA is highly adaptive to data, efficiency estimates based on single measurements are very biased and unreliable if reported without estimating their error distributions.

In limited data settings, Simar and Wilson (1998) proposed a bootstrapping method that estimates a confidence band for the frontier by reassigning efficiency scores across buildings. This resampling across buildings attempts to simulate input stochasticity by reutilizing the same fixed measurements. The implied assumption of this method is that efficiencies are not a function of individual and potentially unobserved building characteristics. Thus, when multiple measurements per building are available, the bootstrapping method is not optimal for estimating uncertainty in energy-efficiency scores; instead, actual repeated measurements of energy consumption should be used.

This paper addresses these various challenges and limitations by introducing a new methodology, SEEF, which combines (1) a hypothesis-testing method to choose efficient frontier shape; (2) a methodology to estimate the uncertainties in the energy-efficiency scores using available repeated measurements; (3) a method to use the estimated uncertainties to create a bias-corrected efficient frontier and a confidence band for the frontier; and (4) a statistical ranking mechanism to compare efficiency scores across the building population.

SEEF Model and Algorithm

SEEF is a stochastic performance-based method to rank a building’s energy efficiency. By introducing SEEF, this paper demonstrates the value of using frontier methods to rank buildings’ energy efficiency and to improve the definition of the energy-efficient frontier. The proposed SEEF methodology extends standard DEA analysis in the field of energy-efficiency scores by incorporating uncertainty into the estimation of the frontier and by introducing a data-driven method to support the choice of frontier shape. SEEF consists of three stages: (1) set up the model by choosing the frontier shape and forming the LPP; (2) run Algorithm 1 to estimate efficiency scores and their confidence intervals; and (3) benchmark buildings by comparing their ranks. Each step is described in the remainder of this section.

Choice of Frontier Shape

To validate the choice of frontier shape, the following hypothesis setup is used Null hypothesis $H_0$: The frontier is a straight line (representing constant returns to scale—CRS). Alternative hypothesis $H_1$: The frontier is piecewise and linearly convex (representing variable returns to scale—VRS).

The following test statistic is used to reject or accept the null hypothesis $H_0$:

$$R = \frac{\sum_{k=1}^{K} \theta_k^{CRS}}{\sum_{k=1}^{K} \theta_k^{VRS}}$$

where $\theta_k^{CRS}$ is the efficiency estimate of building $k$ as calculated with respect to a CRS frontier, and similarly for VRS. If $H_0$ is true, then $R$ will be close to 1; otherwise, $R < 1$. Because the a priori distribution of $R$ is unknown, its empirical distribution is determined from the data set (see Algorithm 1 for details). The empirical distribution of $R$ is used to define a critical value for the null hypothesis and against which to check the statistic $R$.

Stochastic Energy Efficiency Frontier Algorithm

Once the frontier shape is known, efficiency scores are calculated by solving LPP 1 ($S, X$). Before explaining the SEEF method, the following assumptions are established:

- Energy consumption is stochastic and best explained by a random variable whose variance is (partially) explained using variables such as climate zone and type of building (Kavousian et al. 2013).
- Service level values are deterministic. For example, the floor area of a house and the size of a family are not stochastic.
- The building sample adequately represents the population of interest; i.e., the sample selection error is not significant.

The three assumptions imply that energy-efficiency scores are also random variables and not deterministic values. Furthermore, the only source of uncertainty in the efficiency score is the stochasticity of energy consumption measurements. As a result, using just a single efficiency value to compare the energy efficiency of two or more buildings is not possible. Algorithm 1 explains how uncertainty of energy consumption is incorporated into the efficiency estimates.

Algorithm 1. Stochastic Energy Efficiency Frontier: Estimating the Bias-corrected and Confidence Interval for Efficiency Scores Based on Repeated Measurements

Repeat for $b = 1, 2, \ldots, B$ (recommended $B > 1,000$)

For each building, calculate $x_b^k$ for $k = 1, 2, \ldots, K$ as follows:

Draw a random value from the empirical cumulative distribution function (ECDF) of energy consumption measurements (Fig. 3):

1. Form the ECDF.
2. Draw a random number from $U[0, 1]$.
3. Interpolate between the two measurements that correspond to the ECDF value closest to the random number.

$$\text{if } \frac{i}{n} \leq u < \frac{i + 1}{n} \text{ then } x_b^k = n \left[ x_i \left( \frac{u - i}{n} \right) + x_{i+1} \left( \frac{i + 1}{n} - u \right) \right]$$

where $n$ is the number of measurements. For the definition of other variables, see Fig. 3.
Use $LPP_1(S, X)$ to calculate efficiency scores for each building. For the case of one input (e.g., electricity consumption) and one service level measure with a finite number of values $(j = 1, 2, \ldots, J)$, the solution is

$$\theta^j_i = \frac{x_i^j}{x_i^k} \quad \text{where} \quad x_i^b = \min_k \{x_k^b; S(x_k^b) = S_j\}$$

$S_j$ is the $j$th LOS observed in the population and $S(x_k^b)$ is the service level of building $k$.

Using the efficiency values of each iteration ($\theta^1_i, \theta^2_i, \ldots, \theta^K_i$), calculate the bias-corrected efficiency score

$$\tilde{\theta}_i = E[\theta^j_i], \quad I = 1, 2, \ldots, K; \quad b = 1, 2, \ldots, B$$

Confidence interval for efficiency scores

$$\tilde{\theta}_i \pm sd(\tilde{\theta}_i) \quad I = 1, 2, \ldots, K; \quad b = 1, 2, \ldots, B$$

To illustrate the algorithm using an example, consider two buildings A and B, each with 90 days of energy consumption data. These 90 values are presented using a random variable distribution; i.e., each is a realization of the energy consumption random variable. Randomly selecting a consumption value for each user is first done. However, instead of directly selecting one of the 90 values, an ECDF of observed consumption values for each user is formed, and then a probability on the ECDF is randomly selected. Finally, the corresponding consumption value is found by interpolating between observed values (Fig. 3). Once the energy consumption values of the two buildings are selected, the energy-efficient frontier is formed. In this case, the building with lower energy consumption compared with the LOS ratio defines the frontier. After specifying the efficient frontier, the efficiency values of each building is found using the formula presented in Fig. 1. This process should be repeated several times (for the purpose of this example, assume 1,000 iterations), and the efficiency values at each iteration must be recorded. At the end of the iterations, the efficiency scores for each building form a distribution (in this case, two distributions exist, each formed using 1,000 values—one for each iteration). The mean of each distribution is the bias-corrected efficiency score, and the 95% confidence interval is the 95% confidence interval of the efficiency score.

**Stochastic Benchmarking of Buildings Using SEEF**

SEEF returns an efficient frontier and an estimate of the uncertainty in the frontier. It forms a distribution of the efficiency score for each building, from which the bias-corrected score and its confidence interval are estimated. The efficiency score distribution is also used to find a stochastic ranking of buildings: in each iteration $b$ of Algorithm 1, record rank $q(b, k)$ of building $k$ among all buildings. Then, form ECDF of $q(b, k)$ to provide a distribution of ranks for building $k$. For example, the ECDF is used to determine a 95% confidence interval for the rank of building $k$. Including reliability in the rankings allows the analyst to make more informed comparisons.

**Experimental Data and Model Setup**

To illustrate the application of SEEF, the method was applied to a 10-min interval smart meter data set of 307 households in the United States, collected between June and September 2010. A smart meter is a device that records the electricity consumption of a building in regular intervals (every 10 min in this case) and communicates that information to a server, on which it is saved for later retrieval. The consumption data are accompanied by a survey of household information, including demographic features, building characteristics, and appliance ownership. The participants are spread across 164 zip codes and cover six out of eight climate zones in the United States. Climate zones in the data set include cold, hot-dry, hot-humid, marine, mixed-dry, and mixed-humid. The two excluded data sets are the very cold and subarctic climate zones because the former only occurs in a few places within the United States and the latter only occurs in Alaska. Climate zones are defined with reference to Building America definitions and are based on annual precipitation and temperature records (PNNL and ORNL 2010). Building types for 307 subjects in the data set include apartment units, townhouses, and detached buildings. Participants are employees of a technology company in Silicon Valley and voluntarily participated in the experiment. They all paid their own electricity bill. More details on the data collection process is provided in Kavousian et al. (2013) and Houde et al. (2012).

The next section describes how a measure of LOS for each house was defined based on the information available in the survey.

**Choice of LOS**

Quantifying the utility that users receive from consuming energy is easier in some contexts than in others. For example, for a manufacturing plant, utility is defined as the number of products that the plant produces. For a retail store, utility is defined as the number of customers it serves (Phylipsen et al. 1997). However, in a household model, quantifying the utility that a household receives from using energy is still an open question among economists, behavioral economists, sociologists, and psychologists (Lutzenhiser et al. 2010). This paper uses a technical approach (Lutzenhiser et al. 2010) to calculate a measure that is correlated with the service that users receive from electricity consumption. This measure, defined as the aggregate expected consumption of the stock of appliances, is called the Appliance Ownership Index (AOI) (Capasso et al. 1994). A household’s AOI is defined as a linear combination of the number of most important appliances, with coefficients of the linear combination being the industry average (or model-based estimates) of the electricity consumption of each piece of appliance. The AOI is normalized (setting the maximum value to one) to form LOS. In
other words, the LOS of user $u$, denoted by LOS$_u$, is calculated using the following formula:

$$\text{AOL}_u = \sum_{i=1}^{k} A_k \times N_{k,u}$$

$$\text{LOS}_u = \frac{\text{AOL}_u}{\text{Max(AOI)}}$$

where $A_k$ is the average electricity consumption for appliance $k$ and $N_{k,u}$ is the number of appliance $k$ that user $u$ owns.

Choice of LOS is context specific and depends on the purpose of energy-efficiency scores studies. LOS is further discussed later in this paper and a sensitivity analysis is performed to examine the effect of the choice of LOS metric on energy-efficiency scores estimates. First, the subsequent sections demonstrate the power of the SEEF model in informing energy-efficiency scores programs using the AOI as the definition of LOS.

**Weather Normalization**

Some of the uncertainty in the energy consumption of a building is the result of local weather variations. If weather data are available, normalizing consumption values for the weather effect is important. An examination of the relationship between energy consumption and outside temperature in the sample shows a significant correlation for temperature lags of up to 24 h. To remove this effect, a stepwise regression model was fit on outside temperature, including lags of up to 24 h:

$$y_i = a_i + \sum_{j} b_{ij} T_{ij} + \varepsilon_i \quad i = 1, 2, \ldots, K; \quad j = 0, 1, 2, \ldots, 24$$

where $T_{ij}$’s are the outside temperature readings for building $i$ recorded $j$ samples ago.

The estimated coefficients $\hat{b}_{ij}$’s are then used to calculate normalized energy consumption by assigning a reference temperature (65 °F) to all households and calculating their hourly consumption using the following regression model:

$$\tilde{y}_i = a_i + \sum_{j} \hat{b}_{ij} \tilde{T}_{ij} + \varepsilon_i \quad i = 1, 2, \ldots, K; \quad j = 1, 2, \ldots, 24; \quad \tilde{T}_{ij} = 65 F \ \forall \ i, j$$

**SEEF Frontier Shape Choice**

The hypothesis test for the shape of the frontier for this data set suggests a nonlinear frontier ($R^2: 0.64; R_{\text{crit}} = 0.91$). Fig. 4(a) illustrates two innovations of SEEF by comparing a linear frontier with a curved frontier. Existing DEA applications in energy efficiency use the linear frontier (dashed red line); however, as Fig. 4 shows, using a linear frontier results in underestimation of the energy efficiency of high energy users. SEEF offers a method to choose the functional form of the frontier based on measured data. Fig. 4(b) shows the bias-corrected frontier and its confidence intervals. Existing DEA applications in energy efficiency treat energy consumption as a deterministic value or resample values from a fixed set of measurements to simulate fluctuations in daily energy consumption. SEEF offers a method for using repeated measurements of energy consumption to estimate bias-corrected and confidence intervals for the efficient frontier.

Note that the bias-corrected frontier does not completely envelop data points as a deterministic frontier would do because of the stochasticity assumption in estimating a bias-corrected efficient frontier. However, the efficiency values are still strictly between zero and one, as explained here: for any given iteration, a random consumption value for each user is selected and the efficient frontier is formed using the collection of those random consumption values. Therefore, for any given iteration, the frontier envelops all points; thus, the efficiency values calculated at any given iteration are between zero and one. Subsequently, the bias-corrected efficiency values that are the averages of the efficiency values of all iterations are still between zero and one. In contrast, each data point on the plot represents the average consumption value that may or may not be enveloped completely by the bias-corrected frontier. In short, the seeming discrepancy of efficiency values higher than one is only the result of the way the points and the frontiers are illustrated. The actual efficiency values are still between zero and one.
Results and Discussion

SEEF results are used in a variety of applications, including ranking buildings based on their energy efficiency, comparing subpopulations of buildings in terms of their energy efficiency, and identifying irregularities in the energy consumption patterns of buildings compared with the overall population. The following sections provide examples of each analysis.

Rank Comparison

Fig. 5 offers a visual aid to compare the efficiency scores of any two buildings. The smaller the overlap of the two confidence intervals of the buildings, the higher is the degree of confidence that one building is more efficient than the other one. As explained in previous sections, the ECDF of the efficiency scores are used to determine the confidence level of this comparison.

For example, in Fig. 5, the efficiency scores drop sharply after the 15th building. For reference, the first 15 buildings are labeled as the efficient buildings subset (or efficient homes) and, similarly, the rest of the population is called typical buildings. The choice of an efficient set depends on the goals of the analysis of energy-efficiency scores; for instance, the first seven buildings may be labeled as efficient because the confidence intervals of their efficiency values are all higher than 0.6 on the efficiency scale. Identifying the efficient set enables a comparison of the characteristics of its buildings with the rest of the population to infer insights.

For example, in the data set, although 65% of typical buildings resided in a marine climate zone, this ratio was 73% for the efficient buildings subset. Because consumption data for outside temperature were normalized, the higher ratio of marine-climate residences in the efficient homes group suggests the effect of other factors that correlate with a mild climate, such as lower penetration of air conditioners or a design based on natural ventilation.

Newer buildings are more strongly represented in efficient buildings compared with the rest of the population: 40% of efficient homes were built after 2004, whereas only 18% of typical buildings were built after 2004. In contrast, although 45% of typical buildings were built before 1975, this ratio is 33% for efficient homes.

Efficient houses in the data set also illustrated higher penetration of air conditioners compared with the rest of the population. In contrast, some items such as dishwashers, electric heaters, second refrigerators, and second or third air conditioners are less common in efficient houses. One possible reason for this observation is that having high-consumption items such as water heaters and central heaters provides some buildings with an extra allowance for energy consumption because of the manner in which efficiency was defined. In other words, having high-consumption items increases the service level of the building, which enables the building to use more energy when considered efficient. This behavior of the frontier method has two implications for practitioners. First, it suggests that efficiency is a relative term that is defined in the context of the service that a building offers. Frontier methods such as SEEF enable a more balanced view of efficiency by comparing buildings of similar levels of service with one another rather than comparing the entire stock of buildings simultaneously. Second, the effect of the LOS measure on energy-efficiency scores emphasizes the importance of choosing an appropriate LOS measure and performing sensitivity analysis before making judgments about the energy efficiency of a building. This insight is discussed subsequently in this paper.

Another important application of efficiency ranks is to prioritize and plan for energy-efficiency scores programs (i.e., benchmarking). Buildings with lower efficiency scores should be given priority in efficiency programs.

Comparing Energy Efficiency of Buildings in Different Subpopulations

The efficient frontier specifies the best-practice technology achievable by a population of buildings. Therefore, a comparison of efficient frontiers of subpopulations of buildings allows the identification of the subpopulation that achieves lower levels of energy consumption for the same LOS. Such an insight may be used in several applications, including comparing the potential for energy efficiency in different populations of buildings and optimizing design practices to design a building with more energy-efficient features.

Fig. 6 shows an example of such a comparison, for which efficient frontiers for buildings in different climate zones are specified. The order of frontiers from left to right suggests that buildings with marine and hot-humid climate zones in the data set showed the highest efficiency values, followed by mixed-dry and mixed-humid climate zones.

Table 2.

Table 2. Appliance Ownership in Efficient Homes versus the Overall Population

<table>
<thead>
<tr>
<th>Appliance Ownership</th>
<th>Penetration in homes</th>
<th>Penetration in the rest of the population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electric water heaters</td>
<td>27%</td>
<td>13%</td>
</tr>
<tr>
<td>Electric central heater</td>
<td>40%</td>
<td>20%</td>
</tr>
<tr>
<td>Dish washer</td>
<td>66%</td>
<td>92%</td>
</tr>
<tr>
<td>Electric heater</td>
<td>8%</td>
<td>19%</td>
</tr>
<tr>
<td>Electric oven</td>
<td>87%</td>
<td>55%</td>
</tr>
<tr>
<td>Second refrigerator</td>
<td>13%</td>
<td>20%</td>
</tr>
<tr>
<td>More than one AC</td>
<td>6%</td>
<td>14%</td>
</tr>
</tbody>
</table>
Time-of-Use Analysis

Comparing energy-efficiency scores based on whole-day consumption may obscure hourly patterns of consumption in buildings. Time-of-use (TOU) analysis reveals buildings that are efficient during certain hours, such as peak hours. Energy-efficiency scores planners also need this distinction because a building that is inefficient during afternoon hours may benefit from a more efficient air conditioner unit, whereas a building that is highly inefficient during evening hours may need a new, more efficient TV or water heater.

Fig. 7 compares the energy-efficiency scores of buildings when estimated based on consumption during peak hours versus off-peak hours. We define the peak versus off-peak schedule as follows: weekdays from 1 p.m. to 7 p.m. are considered peak hours; weekdays from 10 a.m. to 1 p.m. and weekends from 5 p.m. to 8 p.m. are considered part-peak hours; and the rest of the week’s hours are considered off-peak (Pacific Gas and Electric 2012). Most buildings stay close to the unit slope line, which implies that buildings that are efficient when measured based on their off-peak-hour consumption are also efficient when measured by their peak-hour consumption.

Comparing SEEF with Other Benchmarking Methods

A common method for measuring the energy efficiency of buildings is comparing energy intensity (energy per square foot) (Schipper 1983). The energy-intensity method is a fast and simple way to conduct a preliminary comparison of buildings of similar size and characteristics. However, the method is too simplistic because it assumes that the entire population of buildings (e.g., with their different types, sizes, and occupancy levels) is comparable universally and with only one metric. In frontier methods terms, the energy-intensity method is comparable to a frontier model with square footage as the LOS and a linear frontier. However, research showed that buildings are not linear systems and, as buildings grow in size, the processes through which they consume energy change (Kavousian et al. 2013). For instance, the volume of a house may increase at a different rate than its square footage; alternatively, the penetration of certain equipment such as pools changes as the building’s size changes.

SEEF solves this problem by comparing the energy consumption of buildings of the same service level without making any assumptions about the linearity of buildings’ energy consumption. Thus, when the efficient frontier is nonlinear, the results of the intensity method are different from those of DEA (Fig. 8).

To make the case clearer, efficient buildings as measured using the intensity method were compared with the 15 most efficient homes as measured using SEEF. As Fig. 8 shows, the energy-intensity method has a bias toward buildings with smaller floor area. In contrast, SEEF identifies efficient buildings across the entire range of building sizes and energy consumption values. In this paper, the discrepancy between SEEF and energy-intensity method is the result of the nonlinearity of the efficient frontier.

Indeed, some of the houses that were not identified as efficient by the energy-intensity method featured several energy-efficiency scores attributes. For example, one such house has full insulation including walls, doors, ceiling, basement, and window insulation, along with double-pane windows and a programmable thermostat, and it consumes the lowest amount of energy among its peer houses. Still, the energy-intensity method designated that house as inefficient. This example of a type I error in identifying...
inefficient buildings declared a building inefficient when in fact it does not have significant potential for improvement. Type I errors lead to inefficient allocation of resources when planning energy-efficiency scores programs.

Discussion of LOS

The choice of LOS affects the results of SEEF. For instance, Fig. 9 shows a comparison of the efficiency scores of this paper’s data set when calculated using different LOS definitions. The upper-right part of the plot shows the correlation between efficiency scores of buildings when measured using different LOS’s. The bottom-left part of the plot visually shows this relationship: the horizontal and vertical axes on each plot represent efficiency scores based on one LOS, and the local regression line shows the relationship of the efficiency scores. Finally, the plots on the diameter show histograms of efficiency scores with respect to each LOS definition.

The correlations of efficiency values range from 25% to approximately 63%, suggesting that the efficiency scores of a building when measured with respect to different LOS’s may differ significantly. In other words, different metrics encourage different behaviors. For example, choosing energy per square foot favors large buildings with fewer occupants but indicates very little about how efficiently the occupants use energy. For an energy-efficiency scores planner, choosing the right LOS definition is important to encourage the appropriate behavior. Once the LOS is specified, the data requirements to benchmark the buildings also become clear. The goal of this paper is to develop a novel method to quantify energy efficiency once the LOS is defined. We chose the appliance ownership index as an example to illustrate the capabilities of SEEF. We also demonstrated the importance of the choice of LOS by comparing efficiency scores calculated using different LOS definitions. However, arguably, the choice of LOS is context

![Fig. 9](image-url)
specific and needs to be made based on the goals of the energy-efficiency scores analysis.

Conclusions

Frontier methods quantify the energy efficiency of buildings by forming an efficient frontier (i.e., best-practice technology) and comparing all buildings against that frontier. The efficient frontier specifies the lowest energy consumption observed at any LOS, defined as the utility that users receive from using energy. Because energy consumption fluctuates with time, energy efficiency is a stochastic variable rather than a deterministic value. Thus, one needs to estimate the uncertainty of this stochastic variable when benchmarking building energy efficiency.

This paper developed the SEEF method to find bias-corrected estimates and confidence intervals for the energy efficiency of buildings based on repeated observations of energy consumption. A method to rank buildings based on SEEF results was also presented. This paper showed how the functional form of the frontier affects the results of the analysis and developed a method to choose the functional form based on the data. A method to calculate and compare the stochastic energy-efficiency scores was also developed, and a ranking algorithm based on the stochastic efficiency scores was introduced.

Using a smart meter data set of 307 residential buildings in the United States, this paper demonstrated how SEEF compares the energy efficiency of different populations at different times. To illustrate the method, the AOI metric (a linear combination of the number of appliances, with coefficients proportional to the average consumption of each piece of appliance) was used as the LOS. The results of SEEF (efficiency scores and ranks) are useful for several follow-up analyses, including comparing different subpopulations of buildings through their efficient frontier and by comparing the efficiency scores of buildings in different time-of-use periods to identify irregular behavior.

SEEF performed better than the energy-intensity method (comparing kWh/sq. ft.) for benchmarking energy efficiency of the smart meter data set. Although SEEF identified efficient buildings across the entire spectrum of building sizes and types, the energy-intensity method showed bias toward smaller buildings. The discrepancy between the results of SEEF and the energy-intensity method is attributed to the nonlinearity of buildings; i.e., the same ratio of kWh/sq. ft. could not be used as a benchmark to compare buildings of all sizes and types. SEEF addresses the nonlinearity of buildings by using a nonparametric frontier that is defined at each service level as the benchmark.

SEEF offers an improvement to existing frontier methods to benchmark the energy efficiency of buildings. In sum, these advantages are as follows:

- SEEF recognizes the stochasticity of energy consumption and treats energy efficiency as a random variable. By doing so, SEEF finds bias-corrected efficiency estimates and their respective confidence intervals, which are used to compare the efficiency of different buildings.
- The ranking process identifies the probability distribution of the efficiency ranks of buildings, which is an improvement over the existing methods that simply compare energy-efficiency scores of buildings and deterministically declare one building more efficient than the other.
- SEEF uses actual data (i.e., repeated measurements) to estimate the uncertainty of the efficiency scores.

In short, the contributions of this paper to the field of energy efficiency include the following:

1. Calling for a new way to quantify efficiency based on the input/output relationship;
2. Demonstrating the value of frontier methods in energy-efficiency scores analysis; and
3. Improving the manner in which the efficient frontier is defined by
   - Formalizing a statistical method for choosing the functional form of the efficient frontier;
   - Developing a method to estimate the sensitivity of efficiency estimates to temporal variations in energy consumption using measured data as opposed to resampling methods (i.e., stochastic nonparametric frontier instead of a deterministic one); and
   - Developing the SEEF algorithm to rank buildings based on their energy efficiency given the LOS and the time-series data for energy consumption.

Although this paper presented SEEF as a novel approach to benchmarking a building’s energy efficiency, the definition of LOS is still open to discussion. A possible follow-up study entails examining the results of the energy-efficiency scores ranking using different LOS definitions and comparing the results critically. Because SEEF compares all buildings using the best-practice technology, it is prone to outliers. Thus, another direction for the further development of SEEF is to devise a method to identify and remove outliers from the data set. Developing a robust statistical ranking algorithm for buildings based on SEEF is planned. As the first step, the performance of SEEF on a large data set of energy consumption is being analyzed. The results of this research are capable of improving the manner in which buildings are benchmarked and, ultimately, improving resource allocation for energy-efficiency scores programs.

Acknowledgments

Amir Kavousian was funded through an Advanced Research Projects Agency-Energy (ARPA-E) grant to Stanford University. Ram Rajagopal was supported by the Powell Foundation Fellowship and TomKat Center for Sustainable Energy. The authors would like to thank Professor Martin Fischer, Professor James Sweeney, Dr. June Flora, Dr. Carrie Armel, and Professor Pravin Varaiya for their helpful inputs and comments during this study. The authors would also like to thank Rene Morkos, Taewan Kim, Austin Becker, and Forest Peterson for their feedback on the level of service discussion, and the respected reviewers for their helpful comments on the paper.

References


