

# Benchmarking the energy efficiency of government buildings with data envelopment analysis

Wen-Shing Lee \*

*Department of Energy and Refrigerating Air-Conditioning Engineering, National Taipei University of Technology, Taiwan, ROC*

Received 17 March 2007; received in revised form 8 July 2007; accepted 9 July 2007

---

## Abstract

Constructing an indicator to measure the effectiveness of energy management is important for energy agencies and authorities. This paper uses multiple linear regression method and data envelopment analysis to examine the effectiveness of energy management. First, the regression method using environmental factors is used to calculate the predicted energy usage intensity of each evaluated building. Data envelopment analysis is then employed to calculate overall energy efficiency, using the predicted energy usage intensity as output and the observed energy usage intensity as input. Data envelopment analysis can be further applied to measure the overall energy efficiency in details by examining environmental factors and management factors. Finally, the energy consumption of buildings can be measured to evaluate the effectiveness of energy management. Buildings examined as a case study in this paper are government office buildings in Taiwan. The results show that most of the buildings evaluated to report a higher predicted energy usage intensity have successfully used efficient energy management methods in energy saving.

© 2007 Elsevier B.V. All rights reserved.

*Keywords:* Energy efficiency; Data envelopment analysis; Regression analysis

---

## 1. Introduction

Constructing an indicator to reveal the effect of energy management is important for energy agencies and authorities. There are two major methods to evaluate the performance of energy consumption in a building: simulation method and statistical analysis method.

The simulation method sets up a mathematical model to calculate theoretical energy consumption and makes a comparison between theoretical energy consumption and observed energy consumption in order to evaluate the performance of energy consumption. Federspiel et al. [1] used numerical software to construct the minimum energy usage intensity (EUI) for laboratories and compared this with observed EUI of the evaluated building. Carriere et al. [2] implemented DOE-2 simulation software to study the energy-savings potential of large buildings. The simulation method used factors including the properties of building construction material, the energy efficiency of related energy-consuming

equipment (such as air-conditioners and lights, etc.), and the usage period to calculate the theoretical energy consumption of the building.

The statistical analysis consists of collecting the survey data and comparing one unit with the others. Chung et al. [3] used multiple regression analysis to build a benchmark table by investigating the relationship between EUIs and the explanatory factors. Furthermore, Filippin [4] analyzed the energy efficiency and emissions of greenhouse gases for 15 public school buildings in the central area of Argentina.

On the other hand, the data envelopment analysis (DEA) has been generally used in the performance evaluation for resource usage. Chauhan et al. [5] used DEA approach to determine the efficiencies of farmers with regard to energy use in rice production activities. The study has helped to segregate efficient farmers from inefficient ones, identify wasteful uses of energy from different sources by inefficient farmers and to suggest reasonable savings in energy uses from different sources. The results showed that, on an average, about 11.6% of the total input energy could be saved if the farmers follow the input package recommended by the study. Hu and Wang [6] used DEA to analyze energy efficiencies of 29 administrative regions in China for the period 1995–2002 with a newly introduced index. The results showed that the central area of China has the worst energy

---

\* Correspondence to: 1, Sec. 3, Chung-hsiao E. Rd., Taipei, Taiwan, ROC.  
Tel.: +886 2 2771 2171x3518; fax: +886 2 2731 4919.

E-mail address: [f10911@ntut.edu.tw](mailto:f10911@ntut.edu.tw).

efficiency and its total adjustment of energy consumption amount is over half of China’s total. Hu and Kao [7] used DEA to find the energy-saving target for APEC economies without reducing their maximum potential gross domestic productions in each year. The major finding is that China has the largest energy-saving target up to almost half of its current usage. Önüt and Soner [8] used DEA to evaluate the energy efficiency in 32 five-star hotels, and the results showed that eight hotels are efficient and 24 hotels are inefficient.

The simulation analysis cannot be commonly used for existing buildings, due to the difficulty of collecting building data such as the heat conductivity of walls. Therefore, the statistical method is generally used for benchmarking the energy consumption of buildings. However, it is important to note that the statistical method only gives the overall efficiency of energy consumption that consists of two factors: environmental factors and energy management factors. The environmental factors include weather conditions, occupant intensity, and so on. The energy management factors include equipment efficiency, operating strategy, and so on. Because the environmental factors cannot be controlled by users, an indicator which reveals the effect of energy management is preferable to using the overall efficiency in the performance evaluation of energy consumption. Due to the difficulty of separating the effect of environmental factors from overall energy efficiency, there is little research analysis of the effect of energy management of buildings. Therefore, we use data envelopment analysis to measure the overall energy efficiency in details by examining environmental factors and management factors. After that, the paper evaluates the effect of energy management of government office buildings in Taiwan as a case study.

**2. Method**

After collecting the data of the evaluated units, the benchmarking process takes place and is composed of two steps: first, a regression analysis is used to calculate the predicted EUI of buildings. Then, the DEA utilizing the predicted EUI as the output and the observed EUI as the input is used to calculate overall efficiency, and is further applied to measure the overall energy efficiency in details by examining environmental factors and management factors.

*2.1. Regression model*

We first use the regression analysis to calculate the regression coefficients of significant factors that affect the energy usage of buildings. The investigated independent variables include buildings’ usage information, such as occupant intensity and building area, and outdoor climate conditions, such as dry bulb temperature, hours of rain, and irradiation amount. The typical regression model is as follows [3]:

$$EUI = a + b_1x_1 + b_2x_2 + \dots + b_kx_k, \tag{1}$$

where  $a$  is the intercept;  $b_1, b_2, \dots, b_k$  are the regression coefficients;  $x_1, x_2, \dots, x_k$  are the significant independent variables.

After establishing the regression model, we calculate the predicted EUI of the evaluated units by using the value of the significant factors and the regression coefficients, as shown in Eq. (2).

$$EUI_{pred} = a + b_1x_{1,obser} + b_2x_{2,obser} + \dots + b_kx_{k,obser}, \tag{2}$$

where  $EUI_{pred}$  is the predicted EUI;  $x_{1,obser}, x_{2,obser}, \dots, x_{k,obser}$  are the observed values of the independent variables.

*2.2. Data envelopment analysis*

DEA is known as a mathematical procedure that uses a linear programming technique to assess the efficiencies of decision-making units (DMU) [6]. A non-parametric piecewise frontier, which owns the optimal efficiency over the datasets, is composed of DMUs and is constructed by DEA for a comparative efficiency measurement. Those DMUs that are located at the efficiency frontier are efficient DMUs. These DMUs own the best efficiency among all DMUs and have their maximum outputs generated among all DMUs by taking the minimum level of inputs.

A regression function depicts the features of average values in parametric analysis while an envelope function focuses on extreme values. The concepts used in the parametric and DEA approaches are demonstrated in Fig. 1 where the case of seven DMUs with single inputs and single outputs is considered [5]. The input and output are shown on the  $x$  and  $y$  axes, respectively. The filled rhombuses represent different DMUs in the data set. The dotted line represents the linear regression line in the parametric approach, depicting the trend in the data points. This approach implicitly recognizes all DMUs on or above this line as efficient. In the case of DEA, however, one draws the envelope (or frontier) of the data set by joining the boundary points by straight lines. In Fig. 1, P1, P2, P3 and P4 are the boundary points. The solid line joining these points forms the envelope for the data set. The DMUs lying on the boundary and represented by points P1, P2, P3 and P4 are considered as efficient DMUs.

In parametric analysis, a single regression equation is assumed to be applicable to all DMU. The approach requires the imposition of a specific functional form (i.e., regression equation, production function, etc.) relating the independent

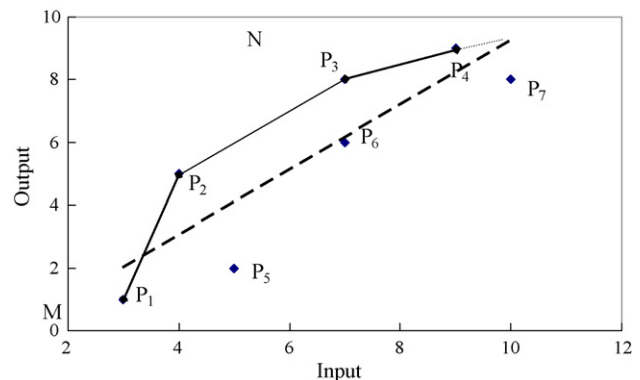


Fig. 1. Comparison of data envelopment analysis and regression analysis [5].

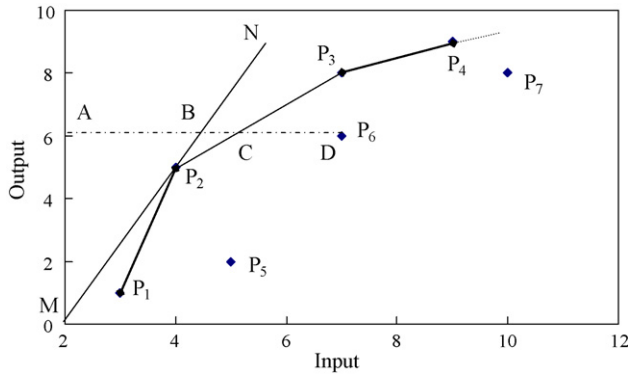


Fig. 2. Relation of input and output from DEA [5].

variables to the dependent variable. In contrast, DEA does not require any assumption about the functional form. It calculates a maximum performance measure for each DMU relative to all other units in the observed population.

A unit can be made efficient either by reducing the input levels and getting the same output (input orientation) or by increasing the output level with the same input level (output orientation) [5]. The input oriented analysis is becoming more common in DEA applications because profitability depends on the efficiency of the operations. In this paper, we adopt an input oriented DEA approach for efficiency estimation.

DEA has two models: constant returns to scale (CRS) DEA model (Charnes et al., [9]) and variable returns to scale (VRS) DEA model (Banker et al., [10]). The CRS model finds each DMU's overall efficiency. The VRS model decomposes overall efficiency into pure technical efficiency and scale efficiency. Overall efficiency is basically a measure by which DMUs are evaluated for their performance relative to other DMUs. However, its value is influenced by scale efficiency, which quantifies the effect of the presence of variable returns to scale in the DMUs. Thus, pure technical efficiency is overall efficiency that has the effect of scale efficiency removed [5].

The concept of these efficiencies is illustrated in Fig. 2 [5]. The line MN represents the envelope of the data set with constant returns to scale. It is a straight line that passes through the origin and the extreme data points. The segment formed by P1, P2, P3 and P4 represent the envelope of the data set with variable returns to scale. The DMU lining on the line MN is consider as efficient and has an overall efficiency equaling to 1; the DMU lining on the segment formed by P1, P2, P3 and P4 has a pure technical efficient equaling to 1. Let us consider DMU P6 in Fig. 2. Its input and output are given by AD and MA, respectively. B and C are the points of intersection of the

line AD with the line MN and the line segment of the envelope of the data set. One can interpret that AB is the ideal input required to produce the output B on MN, if constant returns to scale were to prevail. However, considering variable returns to scale to be a realistic phenomenon, one can relax the input requirement to be equal to AC to be able to produce the output B on MN. One can now define the various efficiencies as follows [5]:

- Overall efficiency = AB/AD;
- Scale efficiency = AB/AC;
- Pure technical efficiency = AC/AD;
- The relationship among these forms of efficiency is given as [5];
- Overall efficiency = [Pure technical efficiency] × [scale efficiency].

### 3. Case study

In this paper, we analyze 47 government office buildings in August and September in Taiwan. Taiwan is an island located between 120° and 122° of east longitude, 22–25° of north latitude. The floor area and the occupants of the evaluated units are provided by their energy manager. Because the electricity is the only energy used in these government office buildings, the electricity usage is the only energy measurement factor and the data is provided from the power utility. The weather conditions, such as outdoor temperature and hours of rain, are provided from 10 climate measurement stations of the Central Weather Bureau. The main data information is shown in Table 1.

### 4. Results and discussion

#### 4.1. Regression analysis

After statistical analysis and using energy usage per floor area as the dependent variable, we find that persons per 100 m<sup>2</sup> of floor area, outdoor temperature and hours of rain as independent variables could build a regression model in which R<sup>2</sup> is bigger than 0.7 (*t* statistics in small parentheses). The regression model is:

$$y = 21.4 + \underset{(9.3)}{8.1} \left( \frac{P - 2.9}{1.7} \right) + \underset{(4.2)}{5.3} \left( \frac{T - 29.8}{0.45} \right) + \underset{(4.6)}{5.7} \left( \frac{R - 22.39}{8.03} \right), \quad (3)$$

Table 1  
The main data information of 47 government office buildings at office time in August and September in Taiwan

	Energy usage intensity (kWh/m <sup>2</sup> 2 month)	Occupant intensity (people/100 m <sup>2</sup> )	Average outdoor temperature (°C)	Average hours of rain (h/month) <sup>a</sup>
Average value	21.43	2.90	29.81	22.39
Standard deviation	10.17	1.70	0.45	8.03
Maximum value	43.19	8.04	30.27	35.50
Minimum value	5.28	0.58	28.67	11.00

<sup>a</sup> Accumulated precipitation more than 0.1 mm/h.

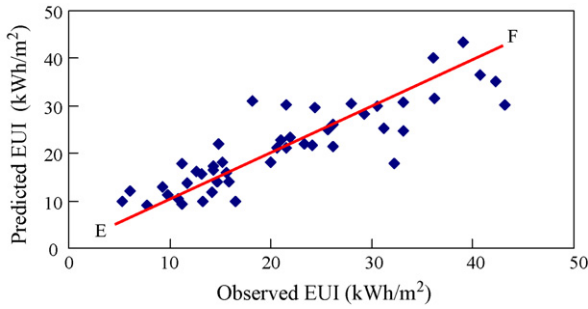


Fig. 3. Plotted distributions between the predicted EUI and the observed EUI of buildings.

where  $R^2 = 0.73$ . Here,  $y$  is the total energy consumption per unit floor area in August and September ( $\text{kWh/m}^2$  2 month),  $P$  the occupant intensity (people/100  $\text{m}^2$ ),  $T$  the average outdoor temperature during office hours, and  $R$  represents the average hours of rain (h/month) during office hours.

The energy consumption of building is composed of energy consumption of air-conditioning, lighting, and power equipment. When it rains, the outdoor air will have a high enthalpy because the relative humidity is about 100%. Because indoor air quality must be maintained at a constant condition, air-conditioning systems have to consume more energy when processing outside air with high enthalpy. This will increase the energy consumption. Therefore, the hours of rain are used as an independent parameter of above regression equation.

We use Eq. (3) to get the predicted EUI of every unit by inputting their occupant intensity, outdoor temperature, and hours of rain. Fig. 3 shows the distribution plot between the predicted EUI and the observed EUI for the observed buildings. The line EF is 1:1 line of predicted EUI to observed EUI. The presence of the over prediction and under prediction of the predicted EUI means that there maybe other factors that exist which affect the observed EUI. Due to the  $R^2$  in the regression analysis is 0.73, it indicates that Eq. (3) already has a good regression relation and prediction of the data.

4.2. Data envelopment analysis

We use the predicted EUI as the output and the observed EUI as the input for DEA analysis in order to calculate overall efficiency, scale efficiency, and pure technical efficiency.

As shown in Fig. 4, the line MN represents the envelope of the data set with constant returns to scale. It is a straight line that passes through the origin and the extreme data point P2 which has the best performance among all evaluated data. The line formed by P1, P2, P3 and P4 represents the envelope of the data set with variable returns to scale. The DMU line on the line MN has an overall efficient equals to 1; the DMU line on the line formed by P1, P2, P3, P4 have a pure technical efficient equals to 1.

The efficiencies of the evaluated unit D is illustrated as an example. Line AB represents the ideal minimum EUI at the output scale (predicted EUI) A in CRS model referring to the ideal minimum EUI can be calculated using the line MN on different output scale. Then, the overall efficiency is calculated

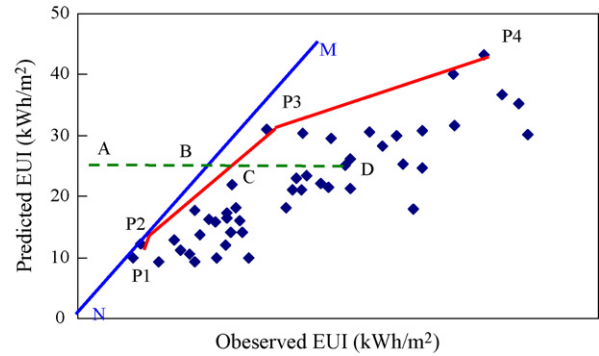


Fig. 4. The efficiency frontier of units.

by dividing AB by AD. In VRS model, it should be noted that line AC is the minimum EUI among all evaluated data on this output scale. Point C, not point B, turns out to report the best performance. Therefore, there exists a scale factor to account for the discrepancy. Accordingly, the overall efficiency can be divided into scale efficiency and pure technical efficiency. The scale efficiency represents the ratio of the ideal minimum EUI (AB) to the minimum EUI (AC) on the output scale and changes with different output scales. Thus, the pure technical efficiency is overall efficiency that has the effect of scale efficiency removed and can be calculated by dividing the real minimum EUI (AC) by the observed EUI (AD).

Fig. 5 shows the relation between scale efficiency and predicted EUI. The scale efficiency decreases as the predicted EUI increases. This is because the predicted EUI is proportional to temperature and the relative humidity of outdoor air, as shown in Eq. (3), a high predicted EUI means the outdoor air has a high temperature and relative humidity. High temperature and high relative humidity will cause the heat rejection system of air-conditioning to work at low efficiency. Thus, the energy consumption will increase. On the other hand, high EUI means the heat rejection per square meter is high, and it is more inefficient when compared with low heat rejection per square meter for a given temperature and relative humidity. Thus, the unit with a higher predicted EUI has a lower scale efficiency.

A higher pure technical efficiency indicates a better effectiveness in energy management. The pure technical efficiency increases as the predicted EUI increases, as shown in Fig. 6, meaning that units have better energy management in high predicted EUI buildings.

The overall efficiency decreases a little as predicted EUI increases, as shown in Fig. 7. The data scatter in Fig. 7 means

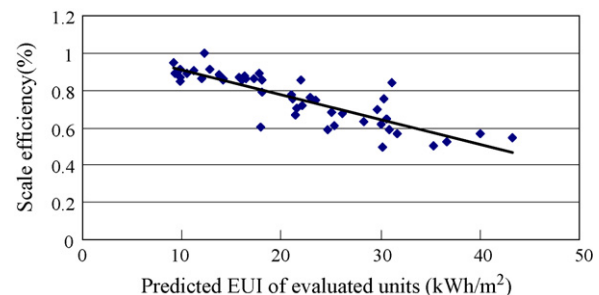


Fig. 5. Relation between scale efficiency and predicted EUI of units.

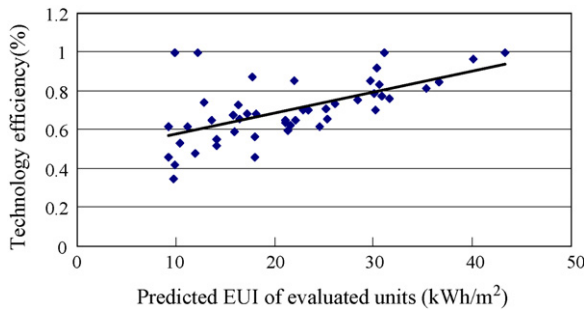


Fig. 6. Relation between pure technical efficiency and observed EUI of units.

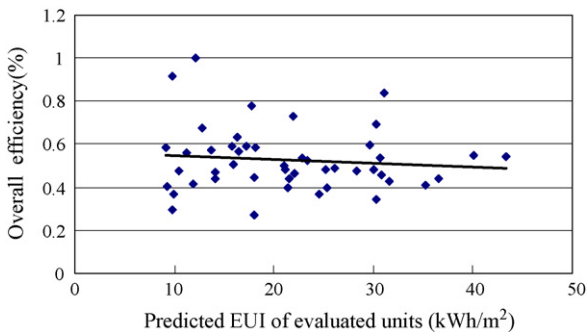


Fig. 7. Relation between overall energy efficiency and observed EUI of units.

the relation between overall efficiency and predicted EUI is not outstanding.

From the above discussion, overall efficiency of building is influenced by environmental factors and energy management, and furthermore the energy efficiency can be divided into scale efficiency and pure technical efficiency. Due to the predicted EUI is calculated by regression method with environmental factors, we deduce that scale efficiency represents the influence of environmental factors in energy consumption. That is, the pure technical efficiency represents the effect of energy management.

## 5. Conclusion

This paper uses multiple linear regression method and data envelopment analysis to find out the effect of energy management. First, a regression method is applied to find out the predicted EUI of evaluated units. Data envelopment

analysis is then employed to calculate overall energy efficiency, using the predicted EUI as output and the observed EUI as input. Data envelopment analysis can be further applied to measure the overall energy efficiency in details by examining environmental factors and management factors. After that, we can evaluate the energy consumption of buildings in the effect of energy management. Finally, the paper evaluates the effect of energy management of government office buildings in Taiwan as a case study. The results show that most of the buildings evaluated to report a higher predicted EUI have successfully used efficient energy management methods in energy saving.

## Acknowledgements

Thanks to the Bureau of Energy, the Ministry of Economic Affairs, and the Taiwan Green Productivity Foundation for data support and budget providing.

## References

- [1] C. Federspiel, Q. Zhang, E. Arens, Model-based benchmarking with application to laboratory buildings, *Energy and Buildings* 34 (2002) 203–214.
- [2] M. Carriere, G.J. Schoenau, R.W. Besant, Investigation of some large building energy conservation opportunities using the DOE-2 model, *Energy Conversion and Management* 40 (1999) 861–872.
- [3] W. Chung, Y.V. Hui, Y. Miu Lam, Benchmarking the energy efficiency of commercial buildings, *Applied Energy* 83 (2006) 1–14.
- [4] C. Filippin, Benchmarking the energy efficiency and greenhouse gases emissions of school buildings in central Argentina, *Building and Environment* 35 (2000) 407–414.
- [5] N.S. Chauhan, P.K.J. Mohapatra, K.P. Pandey, Improving energy productivity in paddy production through benchmarking—An application of data envelopment analysis, *Energy Conversion and Management* 47 (2006) 1063–1085.
- [6] J.-L. Hu, S.-C. Wang, Total-factor energy efficiency of regions in China, *Energy Policy* 34 (2006) 3206–3217.
- [7] J.-L. Hu, C.-H. Kao, Efficient energy-saving targets for APEC economies, *Energy Policy* 35 (2007) 373–382.
- [8] S. Önüt, S. Soner, Energy efficiency assessment for the Antalya region hotels in Turkey, *Energy and Buildings* 38 (2006) 964–971.
- [9] A. Charnes, W.W. Cooper, E. Rhodes, Measuring the efficiency of decision making units, *European Journal of Operational Research* 2 (1978) 429–444.
- [10] R.D. Banker, A. Charnes, W.W. Cooper, Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science* 30 (1984) 1078–1092.