Applied Energy 203 (2017) 454–462

Contents lists available at ScienceDirect

Applied Energy

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

Benchmarking by convex non-parametric least squares with application on the energy performance of office buildings

William Chung *, Iris M.H. Yeung

Department of Management Sciences, City University of Hong Kong, Kowloon Tong, Hong Kong

HIGHLIGHTS

- A novel CNLS-benchmarking system is developed.
- CNLS-benchmarking system has better model fit than OLS-benchmarking system.
- CNLS-benchmarking system does not require any function assumption.
- An illustrative application of Hong Kong office buildings is given.

ARTICLE INFO

Article history:
Received 21 January 2017
Received in revised form 31 May 2017
Accepted 12 June 2017

Keywords:
Benchmarking models
Convex non-parametric least squares
Building energy performance

ABSTRACT

Regression analysis can be used to develop benchmarking systems for the energy performance of office buildings. A linear regression model can be developed using ordinary least squares (OLS) regression analysis to normalize the factors that affect the energy consumption performance of office buildings and develop the benchmarking model. Poor model fit and the assumption of linearity of OLS are the limitations in developing a reliable benchmarking model. In this study, we introduce and discuss the use of convex non-parametric least squares (CNLS) to develop a benchmarking model using the resulting hyperplanes. CNLS is advantageous in that (i) it is a non-parametric regression method, (ii) does not specify the functional form a priori, and (iii) is used to estimate monotonic increasing and convex functions. The resulting benchmarking model can be enhanced with a good model fit using the three advantages. An illustrative application to office buildings is also provided.

1. Introduction

The building sector is widely recognized as a major electricity consumer [1]. Hong et al. [2] mentioned that buildings consume more than one-third of the world's primary energy. In Hong Kong, buildings account for 89% of the total electricity consumption at end-use level.[1] Therefore, promoting energy efficiency in buildings is an effective measure to conserve energy; this initiative is driven by energy policies, such as energy disclosure, rating, benchmarking, and labeling. Consequently, reducing energy use in buildings with energy-efficient technologies becomes feasible.

Energy policies have been introduced to evaluate the energy performance of buildings. For example, Chung [3] considered four mathematical methods to develop benchmarking systems for energy consumption performance; these methods are simple normalization (Simple), ordinary least squares (OLS) regression analysis, stochastic frontier analysis (SFA), and data envelopment analysis (DEA).

Borgstein et al. [4] comprehensively reviewed the available methods for non-domestic buildings. They considered five kinds of methodologies: engineering calculations, simulation, statistical methods, machine learning, and other methods.

Apart from Simple, OLS, SFA, and DEA mentioned in [3], Borgstein et al. [4] also reviewed change-point regression, Gaussian process, and Gaussian mixture regression discussed in [5], TOPSIS in [6,7], and correction factors in [8]. Recently, Capozzoli et al. [9] used linear mixed effect model to build a linear model for benchmarking the energy performance of 100 out-patient healthcare centers.

Chung [3] presented additional descriptions of the abovementioned mathematical methods (Simple, OLS, SFA, and DEA); the author mentioned that the benchmarking models can be developed

* Corresponding author.
E-mail addresses: william.chung@cityu.edu.hk (W. Chung), msiris@cityu.edu.hk (I.M.H. Yeung).

[1] http://www.epd.gov.hk/epd/english/climate_change/bldg.html.

http://dx.doi.org/10.1016/j.apenergy.2017.06.023
0306-2619/

for different applications, and explained how to conduct benchmarking processes.

The selection of the most appropriate analysis method depends on the accuracy, sensitivity, versatility, speed, cost, reproducibility, and ease of use [10]. EPA [11] considered that OLS is a technically rigorous approach and yields descriptive linear equations that are statistically valid and easily replicable.

Chung et al. [12] developed a linear regression model using OLS called the OLS-benchmarking system; they used this model to normalize the factors affecting energy consumption performance and to develop the benchmarking model. The linear regression model may incur a poor model fit (e.g., small R^2 *coefficient* of determination). Braun et al. [13] squared variables to convert the non-linear regression model into an accurate linear regression model. Therefore, statistical benchmarking can effectively identify the energy performance level of a building; however, accuracy is a key issue in applying statistical method. Borgstein et al. [4] presented the same observation.

Therefore, other regression methods, such as convex non-parametric least squares (CNLS), must be considered for an improved model fit.

Kuosmanen [14] derived the representation theorem for CNLS, which is a non-parametric regression method. Since then, CNLS has attracted considerable interest in the literature of productivity efficiency analysis [15]. Several empirical applications have been reported in various areas, such as power generation [16] and electricity distribution [17]. CNLS avoids the functional form assumption and obtains a better model fit compared with OLS.

Kuosmanen and Kortelainen [18] described the use of CNLS to determine the inefficiency estimation by two stages. The first stage conducts CNLS estimation to find all the CNLS residuals. On the basis of the CNLS residuals, the second stage follows the classic SFA study by Aigner et al. [19] to disentangle the half-normal inefficiency from noise. However, all inefficiency terms equal to zero when we attempt to follow this inefficiency estimation to calculate inefficient office buildings (Section 3.2.1). This finding may be due to the non-negatively skewed resulting CNLS residuals.

Hence, we consider another approach, similar to the OLS-benchmarking system in [12], to utilize CNLS estimates and residuals in building a benchmarking system for an improved model fit. However, we find that CNLS cannot be directly used to develop a benchmark table, and the resulting regression models may affect the benchmarking process as in the OLS-benchmarking system. Obviously, a knowledge gap exists between the use of CNLS estimations and the OLS-benchmarking system.

To fill the gap and resolve the aforementioned difficulties, we propose a CNLS-benchmarking system, which combines the approach of the OLS-benchmarking system and CNLS estimations. The contributions of the study are as follows. (i) We develop an approach for using CNLS to build a benchmarking system, such as the one in [12]. (ii) We show that CNLS presents better R^2 coefficient of determination compared with OLS; this feature is critical for developing a benchmark table when the R^2 of the resulting regression model by OLS is small. (iii) We propose a method to calculate the benchmarking score of non-reference buildings for conducting the benchmarking process.

The rest of the paper is organized as follows. Section 2 provides the background of the OLS-benchmarking system. Section 3 discusses why the CNLS estimation cannot be directly used in the OLS-benchmarking system and how to transform the OLS-benchmarking system into the CNLS-benchmarking system with CNLS estimation. Section 4 provides a case study of the benchmarking energy efficiency of Hong Kong office buildings using both systems. Finally, Section 5 elaborates the conclusions of the study.

2. OLS-benchmarking system

The following steps based on [3,12] are used to develop a regression-based benchmarking system after the collection of data from a set of reference office buildings.

Step 1 [Data normalization]: A simple adjustment of the observed performance (e.g., climate adjustment of energy efficiency by degree-day normalization) is conducted. We propose to use the climate correction factor to adjust the space heating/cooling consumption to the climate for a normal year. Following the degree-day-adjustment method used by Haas [20], the climate correction factors (CCF) are defined as follows:

$$CCF = P_{cooling}\frac{CDD_{normal}}{CDD} + (1 - P_{cooling}),$$

where

CDD_{normal} = normal cooling degree days (CDD), average of CDD over a certain period, like 20 years, with cut-off temperature 18.3 °C;

CDD = cooling degree days for a specific year; and

$P_{cooling}$ = proportion of space cooling energy consumption to total energy consumption obtained by the survey.

Step 2 [Linear regression model]: The regression model is built to determine the relationship between the adjusted performance and the selected significant factors (e.g., climate-adjusted energy efficiency and several other significant factors corresponding to office building characteristics, such as energy system, building age, and operation hours). In this step, we need to obtain a best-fitted linear regression model from the standardized data in the following form:

$$Y = \alpha + \beta_1 x_1^* + \cdots + \beta_m x_m^* + \varepsilon^{OLS}, \tag{1}$$

where Y is the climate-adjusted energy efficiency performance, $x_1^*, \ldots, x_m^*$ is a set of significant standardized factors, and ε^{OLS} is the random error.

Step 3 [Benchmark table development]: The adjusted performance for the significant factors is normalized to form a benchmark table. The benchmark table is considered a yardstick for buildings to be benchmarked. In particular, the normalized performance Y_{norm} is calculated using the result of Eq. (1) and is given by

$$Y_{norm} = Y_o - \beta_1 x_1^* - \cdots - \beta_m x_m^*, \tag{2}$$

where Y_o is the observed energy efficiency performance (climate adjusted), and $\beta_1, \ldots, \beta_m$ are determined in Eq. (1). Given n reference buildings, a set of Y_{norm}, $\{Y_{norm(1)}, \ldots, Y_{norm(n)}\}$ can be considered a random sample of Y_{norm} from the population. This set of Y_{norm} measurements constitutes the benchmark basis and can be used to form a benchmarking percentile table for ranking purposes. Moreover, $\{Y_{norm(1)}, \ldots, Y_{norm(n)}\}$ can be considered to provide an empirical cumulative distribution function of Y_{norm}, $ECDF_n(Y_{norm})$, if the benchmark percentile table is to be constructed. That is, given an observed random sample $x_1, x_2, \ldots, x_n$, an $ECDF_n(Y_{norm})$ is the fraction of sample observations less than or equal to the value x. More specifically, if $y_1 < y_2 < \cdots < y_n$ are the order statistics of the observed random sample $\{Y_{norm(1)}, \ldots, Y_{norm(n)}\}$, with no two observations being equal, then the empirical distribution function is defined as:

$$ECDF_n(Y_{norm}) = \begin{cases} 0, & for\ Y_{norm} < y_1 \\ k/n, & for\ y_k \leqslant Y_{norm} < y_{k+1} \\ 1, & for\ Y_{norm} \geqslant y_n \end{cases}$$