Energy Economics 64 (2017) 373–383

Contents lists available at ScienceDirect

Energy Economics

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

Spatial statistical methods applied to the 2015 Brazilian energy distribution benchmarking model: Accounting for unobserved determinants of inefficiencies

Guilherme Dôco Roberti Gil [a], Marcelo Azevedo Costa [a,*], Ana Lúcia Miranda Lopes [b], Vinícius Diniz Mayrink [c]

[a] Department of Production Engineering, UFMG, Brazil
[b] Department of Management Sciences, UFMG, Brazil
[c] Department of Statistics. UFMG, Brazil

ARTICLE INFO

Article history:
Received 13 June 2016
Received in revised form 6 April 2017
Accepted 9 April 2017
Available online 28 April 2017

Keywords:
Data envelopment analysis
Second stage analysis
Spatial statistics
Bayesian analysis

ABSTRACT

In 2015 the Brazilian regulator presented a DEA benchmarking model to set the regulatory operational cost goals, to be reached in four years for 61 electricity distribution utilities. The DEA model uses: adjusted operational cost as the input variable, seven output variables and weight restrictions. Although non-discretionary variables or environmental variables are available in the dataset, the regulator argued that no statistically significant correlation was found between the DEA efficiency scores and the non-discretionary variables. This study evaluates the statistical correlation between the DEA efficiency scores and the available environmental variables. Spatial statistic methods are used to show that the efficiency scores are geographically correlated. Furthermore, due to Brazil's environmental diversity and large territory it is unlikely that only one environmental component is sufficient to adjust inefficiencies across the Brazilian territory. Thus, a new combined environmental variable is proposed. Finally, a second stage model using the proposed environmental variable and accounting for a spatial latent structure is presented. Results show major differences between original and corrected efficiency scores, mainly for utilities located in harsh environments and which originally achieved lower efficiency scores.

1. Introduction

The most commonly used benchmarking models in electricity distribution regulation are: Data Envelopment Analysis (DEA; Charnes et al., 1978), Stochastic Frontier Analysis (SFA; Aigner et al., 1977; Meeusen and van den Broeck, 1977), Corrected Ordinary Least Squares (COLS) (Richmond, 1974) and Stochastic Semi-nonparametric Envelopment of Data (StoNED; Kuosmanen, 2006; Kuosmanen and Kortelainen, 2012). Briefly, DEA is a non-parametric linear programming model proposed by Charnes et al. (1978) which creates the efficiency frontier using a convex linear combination of inputs and outputs of decision making units (DMUs). SFA requires a parametric equation of the efficiency frontier and assumes a compound error, which represents deviations from the frontier. The compound error is the sum of stochastic inefficiencies and stochastic noise. StoNED is similar to SFA and DEA, with a compound stochastic error and with a non-parametric, piecewise linear frontier. Lopes and Mesquita (2015) have shown that these models are very popular among the European electricity distribution regulators.

In general, input and output variables used in DEA, SFA, COLS and StoNED models are associated with controlled factors, i.e., production variables that can be managed by the decision maker in order to improve efficiency. Another set of variables - not necessarily less important - can affect production and are, generally, non-manageable. These variables are known as environmental or contextual variables (Ray, 1988).Examples of contextual variables are climatic factors (Yu et al., 2009) such as temperature, precipitation; soil type, farmers' level of education (Ray and Ghose, 2014); among others. The environmental variables affect the efficiency of companies but are, generally, beyond the scope of company's decisions.

Many alternatives have been proposed to adjust efficiency using environmental factors, such as one stage or second stage analysis. Benchmarking models such as SFA and StoNED allow the inclusion of environmental variables with the input and output variables, using one stage. If the efficiencies of DMUs are estimated using DEA, then second stage analysis is the most common approach. Second stage is based on regression models in which independent variables are the environmental variables.

The analysis of environmental variables was first introduced in DEA models by Banker and Morey (1986), which included the environmental variables in the model as a regular input/output variable. Ray (1988) introduced the second stage analysis, i.e., the efficiency scores are first

* Corresponding author at: Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627 – Pampulha, Belo Horizonte, MG ZIP 31270-901, Brazil.
E-mail address: macosta@ufmg.br (M.A. Costa).

http://dx.doi.org/10.1016/j.eneco.2017.04.009
0140-9883/

estimated using the DEA model and then are correlated to the environmental variables. Ray (1991) included linear regression modeling to evaluate the statistical significance between the efficiency scores and the environmental variables. Since the efficiency scores are within the range 0–1, different statistical regression models such as Tobit regression (Tobin, 1958), maximum likelihood models (Aigner et al., 1977), Truncated regression (Johnson and Kuosmanen, 2012), ordinary least squares (OLS) (Montgomery et al., 2012), among others, can be applied. Simar and Wilson (2007) propose a Data Generating Process (DPG) using a truncated regression and bootstrap procedure to provide valid statistical inference in second stage. The second stage analysis is useful to assist management decisions: the impact of significant environmental factors that negatively affect productivity can be minimized. For example, Ray and Ghose (2014) identified that farmers with higher levels of education and greater access to cutting-edge technologies have better productivity scores. Therefore, public policies could be implemented to increase the levels of farmers' education. Yu et al. (2009) identified statistical significance between weather, cost and quality performance in electricity distribution companies.

It is important to highlight that a critical problem in second stage modeling is the often neglected assumption that the second stage environmental variables do not affect the support of the input and output variables in the first stage. This is known as the separability assumption. Daraio et al. (2010) proposes a non-parametric statistical test for separability. If the separability assumption does not hold, conditional efficiency estimators (Daraio and Simar, 2005, 2007a,b) are appropriate to investigate environmental variables in nonparametric frontier models.

The foundation of second stage analysis is that the estimated efficiency scores using input and output controlled variables can be updated based on the impact of environmental variables. That is, companies located in a favorable environment must have their efficiency scores decreased, in general, since the environment partially contributes to a higher efficiency score. On the contrary, companies located in a harsh environment must have their efficiency scores increased, in general, since the harsh environment prevents the companies from achieving higher efficiency scores. Second stage modeling to adjust efficiency scores are proposed by Simar and Wilson (2007), Banker and Natarajan (2008) and elsewhere. Second stage analysis depends on the nature of the problem being analyzed. If DMUs are subject to environmental settings, it is convenient to use second stage analysis. That seems to be the case for most published studies, including those concerned with regulatory purposes. Furthermore, it can be assumed that the geographic location of DMUs can also be seen as a proxy of the environment, i.e., geographically closer DMUs may be subject to the same environmental setting. This is the foundation of spatial statistical analysis.

In the specific case of Brazilian regulation, second stage analysis may change significantly the efficiency scores of the distribution service operators (DSOs). Brazil is a country as large as a continent with 8.5 million km^2 and it is the 5th largest country in the world with 27 states, most of them larger than some European countries. It covers several climatic zones such as the humid tropics in the north, the semi-arid northeast and temperate areas in the south. These climatic differences lead to major ecological diversity, forming distinct biogeographic zones or biomes: the Amazon Rainforest, the largest tropical rainforest in the world; the Pantanal, the largest floodplain; the Cerrado, savannas and woodlands; the Caatinga semi-arid forests; the fields of the Pampas; and the tropical Atlantic rain forest. For instance, the dry season is very strong in the northeast, in which some municipalities face lack of rain for a few months, or even years. On the contrary, the north, south and southeast of Brazil face critical problems in the raining season like floodings, landslides, etc. Therefore, it is unlikely that the geographic location of the energy distribution companies does not impact their operational costs.

This study applies spatial statistics to evaluate whether estimated 2015 DEA efficiency scores of electricity distribution companies are geographically clustered in the Brazilian territory. A second stage based on stochastic frontier analysis (Aigner et al., 1977) with a latent spatial structure, to account for possible unknown geographical variation of the outputs is proposed. Corrected efficiency scores are estimated using environmental variables and the spatial latent structure. Results show major differences between original and corrected efficiency scores, mainly for companies that originally achieved lower efficiency scores. In addition, the electricity distribution companies located in risky areas, such as areas with flooding, dry regions, or poor regions, have their final efficiency scores increased; whereas electricity distribution companies with higher scores and located in wealthier regions have their final efficiency scores slightly decreased. On average, the new efficiency scores are higher than the original scores.

This paper is organized as follows. Section 2 reviews the second stage analysis for the DEA model and some elements of spatial statistics. It also introduces a new combined environmental analysis and presents the proposed second stage model with non-discretionary and geographically latent variables. Section 3 shows the results. Discussion and conclusion are found in Section 4.

2. Materials and methods

2.1. Background

On June 4, 2014, the Brazilian National Electricity Energy Agency (ANEEL) began a debate with Brazilian society regarding rules and methodologies for defining the revenues of electricity distribution utilities for the 4th Periodic Tariff Review Cycle (4PTRC) through public hearings 023/2014 (AP023). On December 4, 2014, ANEEL presented in Technical Note (TN) 407/2014, the proposed model to calculate regulatory operational costs. The technical note introduces an input oriented, non decreasing return to scale (NDRS), Data Envelopment Analysis - DEA model. This model uses adjusted operational cost as the input variable and seven output variables: high voltage network extension, overhead network extension, underground network extension, weighted power consumption, total number of consumers, estimated number of consumer-hours with interrupted energy, and total amount of non-technical losses (Mega-Watt). The database consists of mean values for the most recent three years, from 2011 to 2013. A total of 61 distribution companies are evaluated, therefore the sample size is $n = 61$. Due to the small data size and the large number of variables, in general, the DEA model generates a larger number of companies with efficiency scores equals to one. To overcome this limitation, weight restrictions are included in the model. Furthermore, non-discretionary variables or environmental variables are available in the dataset. Nevertheless, the TN 407 argues that no statistically significant correlation was found between the efficiency scores and the non-discretionary variables. In addition, it argues that the quality variables defined as the estimated number of customer-hours with interrupted energy and total amount of non-technical losses were able to capture any underlying correlation between efficiency and non-discretionary variables.

Among the contributions, Bogetoft and Lopes (2015) suggest some improvements to ANEEL DEA model. The first suggestion is to include the number of distribution transformers as a new output variable. An extensive simulation study identified this variable as one important output, which is missing in the original model. The second suggestion is to exclude two distribution companies which were identified as outliers. The third suggestion is to evaluate two environmental variables: rain precipitation and frequency of interrupted energy (FEQ) in the second stage. The environmental variables were evaluated using univariate Tobit regression models (Tobin, 1958). Nevertheless, on April 24, 2015, ANEEL presented the final model in Technical Note 66/2015, in which the model presented previously (TN 407/2014) was not changed. That is, the effects of non-discretionary variables were not accounted for in