

DEFINITION OF EFFICIENT REFERENCES FOR REGULATORY BENCHMARKING OF DISTRIBUTION COMPANIES

DEFINICIÓN DE UNA REFERENCIA DE EFICIENCIA PARA LA REGULACIÓN POR COMPARACIÓN DE EMPRESAS DE DISTRIBUCIÓN

Raúl Sanhueza H.¹ Horacio Díaz R.² Ildefonso Harnisch V.³

Recibido 15 de enero de 2008, aceptado 8 de septiembre de 2008

Received: January 15, 2008 Accepted: September 8, 2008

RESUMEN

Se formula un procedimiento basado en el análisis envolvente de datos y la metodología p-mediana para determinar una empresa de referencia eficiente para el proceso de regulación de las empresas de distribución eléctrica. Se determinan las áreas de distribución típicas para establecer las tarifas de distribución en un proceso cuyo núcleo regulador está basado en el concepto de empresa eficiente. Se aplica dentro del marco del último proceso regulador chileno.

Palabras clave: Sistemas de distribución, regulación por comparación, empresa eficiente, análisis envolvente de datos, p-mediana.

ABSTRACT

The data envelopment analysis and the p-median methodology are formulated as tools to determine efficient references for regulatory benchmarking of electrical distribution companies. Distribution typical areas are determined for the purpose of establishing distribution tariffs in a process whose regulatory core is based on the efficient company concept. Its application within the framework of the last Chilean regulatory process is illustrated.

Keywords: Distribution systems, benchmark regulation, efficient company, data envelopment analysis, p-median.

INTRODUCTION

During the last years, important changes have taken place in the electrical sector in many countries around the world. These changes have been carried out both as segmentation and privatization processes. These changes have been possible due to the development of new legal frameworks where the role of the State has been transformed from a producer and company owner agent into a regulatory agent for those segments, such as electrical distribution, that constitute a natural monopoly.

In several countries, a benchmark regulatory scheme has been adopted during the distribution stage, based on the “yard-stick competition” format and efficient company regulation concept. This scheme aims at promoting efficiency and the participation of private stakeholders [3, 4]. Under this scheme, the regulator awards a concession permit to a private company to install and operate networks

dedicated to distribute electricity in a specific geographic area. The right to do business in an area goes along with several obligations, like supply to all that request it, adequate quality of service, etc.

To determine the remuneration for this monopolistic activity, the scheme considers the establishment of a comparison competition, where companies compete with an efficient fictitious company that is called the *model company*. In simple terms, the model company is a ‘mock up’, generally created by consultants, of a company that would supply the electrical demand for the next period in an optimal manner, namely, at a minimum total investment and operating cost.

This scheme, with the competition between the real and the fictitious company, allows the regulator to solve the production efficiency problem through decoupling tariffs from the company’s actual costs. If the fixed tariffs create

¹ Escuela Universitaria de Ingeniería Eléctrica - Electrónica. Universidad de Tarapacá. Casilla 6-D. Arica, Chile. E-mail: sanhueza@uta.cl

² Escuela Universitaria de Ingeniería Eléctrica - Electrónica. Universidad de Tarapacá. Casilla 6-D. Arica, Chile. E-mail: hdiaz@uta.cl

³ Escuela Universitaria de Ingeniería Eléctrica - Electrónica. Universidad de Tarapacá. Casilla 6-D. Arica, Chile. E-mail: harnisch@uta.cl

losses for the real company, it will have to adjust its level of efficiency or to take up such losses. If the real company is able to produce at a lower cost than the model company, it obtains above-normal earnings.

In this process, an important issue to highlight and that has not received enough attention is the definition of the typical distribution areas⁴ and the determination of the reference company. Most commonly, the method used by regulators consists on classifying and grouping distributor companies based on dynamic clustering, trying to identify similar companies to create sets or service areas, and the arbitrarily selection of a real company to represent each category [6, 7]. The inconvenience in this method is that it assumes identical optimized costs for all companies in a same set or, in other words, that a real company with higher costs in a set is more inefficient. In addition, the method does not take in consideration the economies of scale that are accessible to the companies and neither uses efficiency considerations when selecting the representative company. All this implies discrimination among companies, regarding the efficiency target to be fixed.

This article proposes the use of data envelopment analysis, DEA, along with the solution of a p-median problem, to determine the typical distribution areas and the reference company for each. Specifically, in its solution process, the method uses DEA to identify the efficient companies to later form group by solving a-median problem. As a methodology, the DEA has been successfully used in other realms of regulation [8]. This article aims to expand the methodological scope that currently exists regarding this field.

A REVIEW OF TARIFF PROCESSES

In order to regulate by means of the efficient company concept and establish evenly balanced comparisons among different real companies, it is important to clarify which are the production characteristics disparities that could exist among them. These disparities are manifested in factors such as total energy sold, geographical size of the service area, distribution density, etc. These factors are relevant and must be taken into account in the process.

For the first tariff studies three areas types were established: urban, semi-urban and rural areas, defining them based on demand density parameters. The process was new, no previous experience could be used as a reference, so that this very general definition was not questioned and was accepted by the companies. Later, as similar regulatory

processes took place in different countries, regulators learned and modified the procedure for tariff fixation, understanding how sensitive it was for company profits and future performance to have a poor area determination.

To illustrate the learning process in this regulatory path, and the changes that took place in the definition of the typical distribution areas, a review is done of the procedures used by the Chilean regulator during the 1996-2000 and 2000-2004 processes.

1996-2000 Process

The bases for 1996, to define tariffs for 1996-2000, established 5 typical areas that were determined from the parameters and criteria shown in Figure 1:

Parameters:

- C1: Number of inhabitants
- C2: Number of customers
- C3: Number of urban households per square kilometer
- C4: Number of low voltage customers per kilometer of low voltage lines
- C5: Energy sold to regulated customer per number of inhabitants
- C6: Total energy sold per kilometer of high voltage lines
- C7: Total power capacity sold per kilometer of high voltage lines
- C8: Total energy sold per total number of customers.

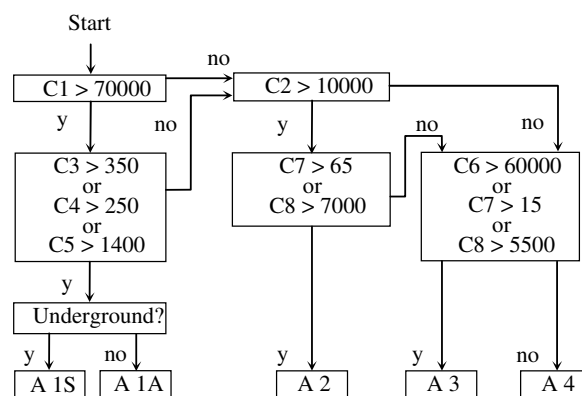


Figure 1. Process to determine typical service areas.

A methodology that made use of a municipality criteria was employed in this process. The procedure considered the design of a single model company that supplied the service zone of a reference real company, and that had the five types of areas. However, this procedure had the inconvenient that the chosen reference real company was too small compared to other regulated companies. That is how, for example, the sample size for the typical rural area was too small compared to other companies that were

⁴ They are defined as "areas where the added value for the distribution activity for each one of those areas is similar one to another".

operating in the same category. The same case happened in the urban area. Consequently, “efficient” costs were established that did not reflect the economies of scale present in larger size companies.

2000-2004 Process

The bases for 2000, to define tariffs for 2000-2004, established six typical distribution areas. The procedure followed to define the areas was based on the cost estimation from real data belonging to the companies. For that purpose, the high and low voltage unit costs were estimated from the adjustment of the following econometric models:

$$\begin{aligned} C_{AT} &= km_{AT} (kW_{AT} + kW_{BT}) e^{(A \ln(km_{AT}(kW_{AT} + kW_{BT})))} \\ C_{BT} &= km_{BT} kW_{BT} e^{(A \ln(km_{BT} kW_{BT}) + B)} \end{aligned} \quad (1)$$

Once the unit costs estimate was known, the standard distribution cost (k\$/kW/year) was determined. It allowed forming company groups so that the costs of companies within each group were not different in a given percentage from the group's average.

The procedure, different to the previous process, introduced the effects of the economies of scale, although it does not consider the relative efficiency of the companies, potentially assigning the same efficient cost to companies with different performances.

As a way to jointly incorporate the economies of scale and the companies' relative efficiencies –issues that must be considered relevant in the determination of distribution areas– a procedure is proposed based on the DEA methodology and the p-median concept, that allows considering both aspects in a unique mathematical scheme.

DATA ENVELOPMENT ANALYSIS

The Data Envelopment Analysis (DEA) [1] is a method of analysis to measure the relative efficiency of a homogeneous number of organizational units that essentially execute the same tasks. In this case the organizational units are electricity distribution companies. Basically, this methodology is centered in determining the most efficient companies in the sample to be used as a benchmark to measure the efficiency of the remaining companies. The most efficient companies are the ones that have no other company or linear combination of companies that produce more of each output (inputs given) or use less of each input (outputs given).

In this manner, if the production technology of the distribution companies is modeled as a correspondence between input-output variables, the proposal for the DEA methodology can be introduced as the following linear programming problem:

$$\begin{aligned} \min \quad & \theta_p \\ \text{sa:} \quad & \theta_p x_{ip} - \sum_{j=1}^n \lambda_j x_{ij} \geq 0 \quad i = 1, 2, \dots, r \\ & -y_{kp} + \sum_{j=1}^n \lambda_j y_{kj} \geq 0 \quad k = 1, 2, \dots, s \\ & \lambda_j \geq 0 \quad j = 1, 2, \dots, n \end{aligned} \quad (2)$$

where: θ_p is the efficiency of the electrical distribution company under evaluation, $x_{1j}, x_{2j}, \dots, x_{rj}$ are the r inputs (for example, investments in distribution substations, among others) and $y_{1j}, y_{2j}, \dots, y_{sj}$ are the s outputs of unit j (for example, energy sold, among others) and $\lambda_1, \lambda_2, \dots, \lambda_n$ are weight factors that enable the convex combination of inputs and outputs in the n distribution companies, respectively.

With the previous model, the efficiency measure may be the result of comparing units of different scale, which may be inadequate. To solve this problem, it is possible to formulate a model that considers the possibility of inefficiencies given the different operative scales of the DMUs (Decision Making Units) [2]. Then, it must incorporate the restriction that all λ_j add one ($\sum \lambda = 1$). This restriction guarantees that the model evaluates pure technical efficiency, without including scale considerations.

$$\begin{aligned} \min \quad & \theta_p \\ \text{sa:} \quad & \theta_p x_{ip} - \sum_{j=1}^n \lambda_j x_{ij} \geq 0 \quad i = 1, 2, \dots, r \\ & -y_{kp} + \sum_{j=1}^n \lambda_j y_{kj} \geq 0 \quad k = 1, 2, \dots, s \\ & \sum_{j=1}^n \lambda_j = 1 \\ & \lambda_j \geq 0 \quad j = 1, 2, \dots, n \end{aligned} \quad (3)$$

The interpretation that can be given to the model is the search of a fictitious distribution company that is a combination of all the companies (with λx inputs and λy outputs) and that produces the same amount of outputs as distribution company p under evaluation, but only using a fraction of the amount of inputs, (θx).

In this manner, it is possible to observe that to determine efficiency, the DEA does not necessarily use a mirror image of reality for the efficient unit, and the usual case is that the unit under evaluation is not compared to a real company, but to a compounded or fictitious company that is a linear combination of other existing companies. This particularity is perfectly consistent with the production theory proposal and assumes the hypothesis of the possibility of using inputs continuously and having convexity in the efficiency frontier.

The set of efficient real units, from whose convex combination results the fictitious efficient unit, is called reference group (*peer group*) and its identification allows planning the improvements of the evaluated unit based on levels that have been actually reached.

SELECTION OF THE REFERENCE UNITS

To identify the reference group, one investigates the value reached by the intensity variables, lambda, in equations (1) and (2). Figure 2 illustrates this condition for the case of four units, characterized by a productive technology that utilizes two inputs and one output. In the figure, the continuous line represents the efficiency frontier considering variable returns to scale, equation (2). Over that frontier line one finds the units found efficient, A and B, which would be efficiency benchmarks for the other units, C and D.

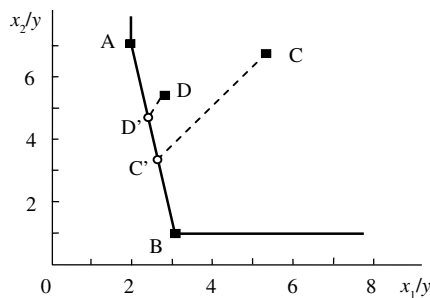


Figure 2. Efficiency characterization.

In figure 2, unit C is inefficient and its projection over the frontier, in direction to the origin, C', represents a fictitious unit that is formed by the convex combination of efficient units A and B. This fictitious unit is the reference unit for the measure of efficiency of unit C, equation (2). Unit B, being nearer to unit C', is the one that has more influence and weight in its formation. Intensity variable lambda associated to B is greater than that associated to A. Thus, given the intensity variable criteria, unit B is the efficiency reference for unit C.

However, this method presents difficulties when the unit considered more influential has little similarity with the inefficient unit. In figure 2, the method designates B as reference to C, although unit A has more similarity to C. Another criterion that could be used to select the reference unit is that indicated by Tulkens [4], which utilizes the dominant unit idea, originally defined by Koopmans [5]. The definition for the dominant unit is: a unit that utilizes several inputs to produce several outputs (products) is technically efficient if, and only if, it is impossible to produce more out of any input without producing less of another output or using more of another input. As indicated in [4], identifying a dominant unit gives more credibility to an efficiency measure, as this identifies an observed reference unit, instead of a convex combination for the unit predominance.

However, as observed in figure 2, there are also problems in this case, particularly with unit D. Unit D is inefficient, but is not strictly dominated by efficient units A and B. To produce the same quantity of output, unit D requires less of input x_2 than unit A and less of input x_1 than unit B. Thus, the method could end in a solution where the dominant subset is empty.

Proposed p-median model

The p-median model is utilized to locate p supply installations in a predefined set of n locations ($n > p$), to minimize the total distance of m demand points to their closer supply points. The standard p-median model [11] is modified to find an alternative solution to the problem of defining distribution service areas.

$$\begin{aligned}
 \text{Min } Z &= \sum_{i=1}^m \sum_{j=1}^n d_{ij} x_{ij} \\
 \text{sa : } & \\
 \sum_{j=1}^n x_{ij} &= 1 \quad i = 1, 2, \dots, m \\
 x_{ij} &\leq x_{jj} \quad i = 1, 2, \dots, m \quad j = 1, 2, \dots, n \\
 \sum_{j=1}^n x_{jj} &= p \\
 x_{jj} &= 1 \quad \forall j \in E_j \\
 x_{ij} &= \{0, 1\} \quad i = 1, 2, \dots, m \quad j = 1, 2, \dots, n
 \end{aligned} \quad (4)$$

where:

- i : index of units
- m : total number of units
- j : index of potential reference units
- n : total number of potential reference units

- E_j : subset of potential reference units that exist and that can not be changes
 d_{ij} : distance between unit i and potential reference unit j
 x_{ij} : variable that is equal to 1 if unit i is assigned to potential reference unit j , 0 if not

In the model, d_{ij} , the distance between unit i and potential reference unit j , is determined considering the inputs and outputs as dimensions, that is:

$$d_{ij} = \sqrt{\sum_{p=1}^r \frac{(x_{pi} - x_{pj})^2}{\bar{x}_p} + \sum_{q=1}^s \frac{(y_{qi} - y_{qj})^2}{\bar{y}_q}} \quad (5)$$

where r is the number of inputs and s the number of outputs of the units; \bar{x}_p and \bar{y}_q are the average of inputs and outputs of the units.

MODEL TO DETERMINE SERVICE AREAS

As a manner to establish the mechanism to define the service areas, production indicators are identified according to the typical nomenclature used in frontier methods, namely, as input variables for resources and output variables for products. For example, it is possible to use the following variables that the authors considered relevant for the real system described in the application:

Output variables

Output variables in distribution are the variables that are fixed in the short term and that efficiently describe the service, the system, and up to a certain point, the environment faced by the distribution companies. The output variables used are:

Energy sold: Energy sales as the primary activity of each company.

Coincident peak maximum power: Maximum power is a *proxy* of the transformation capacity required by the distribution company to allow delivery of energy to customers at peak demand hours. This reflects the fact that the distribution system must be designed to face peak demand occurrences, even if they are high above the mean demand. Considering peak power ensures that the distribution companies that need higher inputs, to face relatively high demands, will not be penalized in the efficiency evaluation.

Input variables

Input variables include costs incurred by distribution companies. The input variables are:

Replacement new value: The replacement new value, RNV, represents the costs of the goods that are needed in the distribution chain, valued at present replacement costs, as reported by the distribution companies to the regulator.

Length of the distribution network: The size of the network is measured by the total number of line kilometers. This output captures the size of the distribution system managed by the company and it ensures that, for example, an extended rural distribution company will not be penalized in the efficiency evaluation compared to a distribution company that renders services only in the city. These variables, together with the number of transformers can also be used to represent the capital cost.

APPLICATION

The methodology is applied using the data base of 35 Chilean distribution companies, as reported to the Superintendence of Electricity and Fuels (SEC) for the tariff fixation process of the year 2000-2004. The RNV used corresponds to that of December 31, 1999.

The DEA model that was used considers variable returns to scale [2], with variables specified before. Table 1 provides results for each company as it relates to the reference groups, unit with greater participation, dominant unit and p-median unit. Companies are represented by a correlative number given by SEC. According to the DEA model, the companies that are efficient are 8, 10, 11, 12, 14 and 15. The relevance of those companies in the reference groups are given by the value of λ shown in columns 2 to 7 in table 1, while columns 8 to 10 show the companies selected according to the greater participation [3], the dominant company [4] and the proposed p-median method. Efficient companies do not have a reference for comparison within the universe being studied.

From the results one can conclude that in most cases the three methods agree on the chosen reference company. However, there are particular cases, such as that of company 26, where the minimum distance method determines company 15 as reference which, given intensity factors and dominance, contributes little to the fictitious company. Something similar occurs with companies 32, 33 and 34, where the minimum distance method determines

company 12 as their reference, although according to the intensity factors and dominance methods corresponds to company 8. Nevertheless, both companies 8 and 12 are very close according to distance; any could be defined as reference.

Table 1. Most relevant company in the reference group.

Cmp.	Reference group						Reference		
	8	10	11	12	14	15	Part.	Dom.	p-median
1		0.01	0.06	0.93			12	12	12
2		0.00	0.18	0.48	0.34		12	14	14
3		0.02	0.18	0.21	0.59		14	14	14
4		0.00	0.26	0.74			12	12	12
5			0.60	0.40			11	11	11
6		0.05	0.95				11	11	11
7		0.04	0.22	0.43	0.31		12	14	11
8	1.00						8	8	8
9			0.02	0.98			12	12	12
10		1.00					10	10	10
11			1.00				11	11	11
12				1.00			12	12	12
13	0.54			0.30		0.16	8	8	8
14					1.00		14	14	14
15						1.00	15	14	15
16				0.56	0.02	0.42	12	12	12
17			0.58	0.42			11	11	11
18		0.26	0.74				11	11	11
19			0.01	0.99			12	12	12
20			0.41	0.59			12	12	11
21			0.81	0.19			11	11	11
22		0.00	0.03	0.97			12	12	12
23		0.02	0.02	0.96			12	12	12
24			0.01	0.99			12	12	12
25			0.02	0.98			12	12	12
26		0.00		0.97		0.03	12	12	15
27			0.03	0.97			12	12	12
28			0.02	0.98			12	12	12
29			0.02	0.98			12	12	12
30			0.02	0.98			12	12	12
31			0.04	0.96			12	12	12
32	0.66			0.34			8	8	12
33	0.62			0.38			8	8	12
34	0.67			0.33			8	8	12
35			0.01	0.99			12	12	12

Finally, table 2 compares the typical service areas obtained by the Chilean regulator for the 2000-2004 process and with the proposed method. The differences arise essentially given the absence of efficiency criteria in the method used by the regulator to build the typical areas.

CONCLUSIONS

The data envelopment analysis methodology is used in this work to determine the typical distribution areas applicable to the regulatory scheme through the efficient company model.

From the data available from distribution companies, the functioning of the activity is characterized through and input – output correlation. Next, the DEA methodology with the VRS model is used to obtain the efficient reference groups of each one of the companies and to establish their belonging to a respective area through a criterion of magnitude of their contribution to the fictitious company.

The use of more information by the companies and, in addition, the use of the DEA methodology give more transparency to the determination of areas, allowing to consider issues such as economies of scale and efficiency in the determination of the benchmark or reference company.

Table 2. Reference companies and area definitions.

DEA p-median method		Regulator method	
Ref.	Companies by area	Ref.	Companies by area
10	10	10	10
11	11, 5, 6, 7, 17, 18, 20, 21,	18	18, 11, 4, 14, 3, 7, 2
14	14, 2, 3,	5	5, 17, 1, 6, 29, 12, 25, 15, 16, 13
12	12, 1, 4, 9, 16, 19, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35	17	17, 28, 23, 24, 27
15	15, 26,	31	31, 22, 8, 21, 9, 32, 35, 39, 36, 26,
8	8, 13	33	33, 30, 34

ACKNOWLEDGEMENTS

The authors acknowledge the support from Universidad de Tarapacá, Project 8723-06.

REFERENCES

- [1] A. Charnes, W.W. Cooper and E. Rhodes. "Measuring the efficiency of decision making units". *European Journal of Operational Research*. Vol. 2, pp. 429-444. 1978.
- [2] R.D. Banker, A. Charnes and W.W. Cooper. "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis". *Management Science*. Vol. 30 N° 9, pp. 1078-1092, 1984.

- [3] S.A.C. Kittelsen and R.F. Førsund. "Efficiency Analysis of Norwegian District Courts". *Journal of Productivity Analysis*. Vol. 3, pp. 277-306, 1992.
- [4] H. Tulkens. "On FDH Efficiency Analysis: Some Methodological Issues and Application to Retail Banking, Courts and Urban Transit". *Journal of Productivity Analysis*. Vol. 4, pp. 183-210. 1993.
- [5] T.C. Koopmans. "An analysis of production as an efficient combination of activities". In: Koopmans, T.C. *Activity analysis of production and allocation*, Cowles Commission for Research in Economics, Monograph. N° 13. NY, JWS. 1951.
- [6] R. Bitu and P. Born. "Tarifas de Energía Eléctrica: Aspectos Conceptuales y Metodológicos". Olade. 1993.
- [7] H. Rudnick and R. Raineri. "Chilean distribution tariffs: incentive regulation". Chapter in (De) *Regulation and Competition: The Electric Industry in Chile*. Ilades-Georgetown University, pp. 223-257. 1997.
- [8] H. Rudnick and J. Donoso. "Integration of price cap and yardstick competition schemes in electrical distribution regulation". *IEEE Transactions on Power Systems*. Vol. 16 N° 4, pp. 939-945. November 2001.
- [9] T. Jamasb and M. Pollitt. "Benchmarking and Regulation of Electricity Distribution and Transmission Utilities: Lessons from International Experience". DAE Working Paper, Department of Applied Economics. University of Cambridge. 2001.
- [10] IPART. "Efficiency and Benchmarking Study of the NSW Distribution Businesses". Commissioned by the Independent Pricing and Regulatory Tribunal of New South Wales, Prepared by London Economics. Sydney. 1999.
- [11] M. Daskin. "Network and Discrete Location: Models, Algorithms and Applications". Wiley Interscience Series in Discrete Mathematics and Optimization. Toronto. 1995.