Consistency Conditions: Efficiency Measures for the Electricity Distribution Sector in South America

Martín Rodríguez Pardina, Martín Rossi
and Christian Ruzzier

Working Paper N° 5
ISBN 987-519-038-1
(May 1999)

CEER
Centro de Estudios Económicos de la Regulación
Instituto de Economía, Universidad Argentina de la Empresa
Chile 1142, 1° piso
(1098) Buenos Aires, Argentina
Teléfono: 54-11-43797693
Fax: 54-11-43797588

(Please see the back pages of this paper for a listing of current working paper titles and subscription ordering information).

The Centro de Estudios de Economía de la Regulación (CEER), is an organization engaged in the analysis of the public regulation of utilities. The CEER is supported by the World Bank, the Regulatory Agencies of the Public Utilities of Telecomunicacions and Power of Argentina, and the Universidad Argentina de la Empresa (Buenos Aires), where the CEER works.

CEER Working Paper Series
Consistency Conditions: Efficiency Measures for the Electricity Distribution Sector in South America
Martín Rodríguez Pardina, Martín Rossi and Christian Ruzzier
Working Paper N° 5
(May 1999)
JEL N°: C4, L9

Abstract: The main goals of this paper are basically two: to compare the relative efficiency of the firms in the electricity distribution sector in South America, and to perform a consistency analysis on the different approaches usually used to measure efficiency. The estimated model considers a single output (customers) and six variables standing for inputs and environmental characteristics (service area, sales, market structure, mains kilometres of distribution, number of employees and transformer capacity). Information on these variables comes from the CIER database for the accounting year ending in 1994. This model is in line with the previous literature on the subject. It has been found that, in general, the consistency conditions are met.

Affiliations of the Authors:
Martín A. RODRÍGUEZ PARDINA
Centro de Estudios Económicos de Regulación, UADE, Argentina
Martín A. ROSSI
Centro de Estudios Económicos de Regulación, UADE, Argentina
Christian A. RUZZIER
Centro de Estudios Económicos de Regulación, UADE, Argentina

CEER
Centro de Estudios Económicos de la Regulación
Instituto de Economía, Universidad Argentina de la Empresa
Chile 1142, 1° piso
(1098) Buenos Aires, Argentina
Teléfono: 54-1-3797693
Fax: 54-1-3797588

Consistency Conditions: Efficiency Measures for the Electricity Distribution Sector in South America

Martín A. RODRÍGUEZ PARDINA
Centro de Estudios Económicos de Regulación, UADE, Argentina.

Martín A. ROSSI
Centro de Estudios Económicos de Regulación, UADE, Argentina.

Christian A. RUZZIER
Centro de Estudios Económicos de Regulación, UADE, Argentina.

I. INTRODUCTION

The goals of this paper are basically two: (i) to compare the relative efficiency of the firms in the electricity distribution sector in South America, and (ii) to perform a consistency analysis over the different approaches generally used in efficiency measurement.

Following the process initiated by Chile in the early '80s, most countries in South America have undergone deep transformations in their electric industries, which include both restructuring and privatization of the prevailing public monopolies. As a result of these processes, a strong change in the role of government has occurred, leaving its producer and firm owner roles to become a regulator of those activities that constitute natural monopolies (namely transmission and distribution).

In this new regulatory role, the comparison of the relative efficiency of several regional monopolies seems to be a potentially valuable instrument to reduce the asymmetry of information that is involved in the regulator-firm relationship. This fact has been recognized in many of the reform processes in which horizontal break-up of transmission and distribution firms was an important ingredient of the transformations.

However, to be useful in the regulatory process this tool needs two conditions to be satisfied. On the one hand, it requires a broad set of comparable firms and detailed information about them. In this respect CIER's effort to build up a regional database is a fundamental contribution for the development of the electric utilities regulation. But, on the other hand, this availability of data, although a necessary condition, is far from sufficient. One must count on adequate techniques that allow an exhaustive analysis of the available data with reference to an appropriate conceptual framework. Our objective in this paper fits into this criterion, analyzing the different approaches at hand in order to contribute to the development of instruments that provide an efficient regulation of the firms in this sector.

The structure of the work is as follows. Section II compares the different methods of estimation, whereas Section III enumerates the consistency conditions. In Section IV, the theoretical model (to be estimated) is formulated, and the different models found in the literature are reviewed. Section V presents the data and the econometric and mathematical programming estimates, while Section VI analyses the consistency conditions explained in Section III. Finally, in Section VII, conclusions to this work are made.

II. A COMPARISON OF DIFFERENT METHODS OF ESTIMATION

The productive efficiency is the firm's ability to produce an output at minimum cost. To achieve that minimum cost the firm must use its inputs in the most efficient way (technical efficiency) and choose the appropriate input mix given the relative price of its inputs (allocative efficiency). Thus, productive efficiency requires both technical and allocative efficiency. Therefore productive inefficiency will tend to be higher than technical inefficiency. One observation emerging from this is that for a comparison of the different methods of estimation to be correct, it requires that all the methods refer to the same efficiency concept.

During the last thirty years, following the pioneer work of Farrell (1957), at least four approaches have been developed in the quest for relative efficiency measurement (relative to the empirically defined actual best practice). These are the mathematical programming non-parametric approach (the so-called Data Envelopment Analysis, DEA) and three parametric approaches: deterministic parametric frontier (DPF), stochastic parametric frontier (SPF), and, if one has panel data, distribution-free approach (DFA) –which makes no assumptions regarding the inefficiency term distribution. Among other things, these approaches differ in the existence or not of a random error, and in the assumption or not of a functional form (a priori) for the technology.

Broadly speaking, the efficiency frontiers analysis can be classified accordingly to the form in which the frontier is specified and estimated. The specification refers to whether the frontier is constructed from a production function or from a cost function. A production function displays the produced quantities as a function of the inputs employed, whereas a cost function shows the total cost of production as a function of the level of output and the input prices.

When choosing between the estimation of a production function or a cost function, it is important to bear in mind the peculiarities of the sector one is studying. An important feature of the regulated utilities, for instance, is that, in general, the firms are obliged to provide the service at the specified tariffs. Therefore, the firms must meet the demand for their service, and are not able to choose the level of output they will offer. Given the exogeneity of the output levels, the firm maximizes profit simply by minimizing the cost of producing a given level of output.

An additional advantage stemming from the use of cost functions, independently of the firms in the sector being regulated, has to do with their flexibility to adapt to situations in which more than one output is produced[1]. Moreover the estimation of production functions allows the measurement only of technical efficiency, but not allocative efficiency, whereas the estimation of cost functions allows for the calculation of both inefficiencies, thus obtaining a measure of the productive (overall) efficiency. Nevertheless, if one wishes to conduct separate estimations on both types of inefficiency it is necessary to make further assumptions.

However, there are other theoretical as well as practical arguments that oppose to the former ones (which favored the estimation of cost functions within this type of industry). Among these is the difficulty to obtain accurate information on input prices. Moreover, the estimation of frontier costs involves the utilization of variables measured in monetary units (data on costs as

[1] Strictly speaking, one could estimate production functions in the case of multiproduct firms using the DEA approach –so the previous analysis applies to the parametric methods only.

well as on input prices is needed), which could be a serious problem if one wishes to make international comparisons. Production functions, instead, only require variables measured in physical units (i.e. homogeneous among countries –or at least more homogeneous).

As a theoretical argument, one could add that whenever there exists public ownership, the firms, in general, will not seek profit maximization as their main goal. Besides, in this kind of firms, prices may not be available nor reliable (Charnes, Cooper and Rhodes, 1978).

From the point of view of estimation, both cost functions and production functions can be estimated with statistical or mathematical programming tools. Non-statistical methods (like DEA) use linear programming techniques. In their usual form, firms are considered efficient if there are no other firms, or linear combination of firms, which produce more of each output (given the inputs) or use less of each input (given the outputs). The principal advantage of the non-parametric approach is that no functional form a priori is imposed on the data. A drawback is that it employs only a subset of the available data, while the rest of the observations are ignored. Moreover, DEA estimates the efficient frontier without making any assumption about the distribution of the error term. The estimations, therefore, lack statistical properties, thus rendering impossible the hypothesis testing. The parametric models, in turn, although allowing for hypothesis testing, might label inefficiency something that actually is a mispecification of the model.

An aspect worth noting is that the efficiency measures obtained with DEA can be very sensitive to the number of variables included in the model. As the ratio (number of variables/sample size) grows, the ability of DEA to discriminate among firms is sharply reduced because it becomes more likely that a certain firm will find some set of weights to apply to its outputs and inputs which will make it appear as efficient (Yunos and Hawdon, 1997). That is to say, a lot of firms might be labeled 100% efficient not because they dominate other firms, but just because there are no other firms or combinations of firms against which they can be compared when there are so many dimensions.

Once decided upon the kind of frontier to be estimated (cost or production) and the estimation technique, the next step is to determine whether such frontier is to be considered deterministic (DEA and DPF) or stochastic (SPF). If the activity frontier is deterministic all the firms share the same cost and production function and every discrepancy between the individual firm performance and the frontier are considered due to inefficiency, thus completely ignoring the possibility of a single firm performance being affected not only by inefficiencies in the management of its resources but also by factors absolutely beyond its control (e.g. adverse weather conditions). An additional disadvantage of the deterministic approaches is that they are very sensitive to the presence of outliers. A single outlier (due perhaps to measurement errors) can have deep effects on the estimations. Moreover, this outlier problem cannot be solved just by increasing the sample size, and it bias the efficiency measurement downwards. This bias will be present both in DEA and DPF estimations. In the former methodology it remains unsolved whether this downward bias is larger or smaller than the upward bias stemming from the addition of a new variable to the model.

Estimation of deterministic frontiers (in DPF technique) involves the utilization of a one-sided error term, which implies that it is possible to define accurately the minimum necessary cost to achieve a given level of output. Therefore, the actual cost is simply the least cost plus an inefficiency term (bound to be equal to or greater than zero by definition). Clearly, the assumption behind this is that every external event, which might affect the cost function, are the same (and of equal intensity) for all firms.

Following the works of Aigner, Lovell and Schmidt (1977) and Meeusen and van de Broeck (1977), the so-called stochastic frontiers made their appearance, based on the idea that the deviations from the frontier could not be entirely under the analyzed firm's control. This approach uses a mix of one-sided and two-sided errors; i.e., given an output level, there exists a minimum feasible cost, but this minimum is stochastic and not precise. The idea is that the external events which influence the cost function are normally distributed (the firm being faced to favorable or unfavorable conditions with given likelihood) instead f being constant. Once considered the likelihood of statistical noise, what remains is termed inefficiency. This decomposition is precisely the nature of the moral hazard problem faced by an imperfectly informed regulator. That is the regulator must establish which fraction of the observed differences between the firms' operational costs is due to inefficiency and which to external factors over which the firms have no control[2].

There are two ways of estimating stochastic frontiers: modified ordinary least squares (SPF-MOLS) and maximum likelihood (SPF-ML). Least squares estimators will in general be less efficient than maximum likelihood estimators because the latter incorporate a priori information on the distribution asymmetry of the error term. The efficiency gains from the use of ML instead of MOLS are a function of the degree of skewness in the distribution of the error term, which is a strictly empirical problem. One positive aspect of the SPF-MOLS approach is that the ranking of the firms will always be equal to that of the cost function residuals, regardless of the assumption made about the distribution of the inefficiency term. That is to say that firms with low costs for given input prices, output quantities and other environmental variables will always be ranked as more efficient.

In general the stochastic frontier models are exposed to three serious drawbacks (Schmidt and Sickles, 1984). Firstly, the inefficiency term estimations, although unbiased, are not consistent (their variance never becomes zero, no matter how much is the sample increased), which really poses a problem if one bears in mind that the goal of the work is the estimation of the sample firms' (in)efficiency. Secondly, both model estimation and separation between inefficiency and noise call for specific assumptions to be made about the distribution of either term. The most used distribution for the inefficiency term in the empirical work is the half-normal distribution. This distribution makes the majority of the firms almost completely efficient, though there is no theoretical reason that prevents the inefficiency to be distributed symmetrically (as the error term is usually assumed to be distributed). Finally, it might be incorrect to assume that the inefficiency is independent from the regressors: if a firm knows its efficiency level, this could affect its input choices.

The preceding problems, which appear under SPF methodology, are potentially solvable using panel data (DFA). The major advantage of using panel data consists in that it allows a greater flexibility in the model build-up, with no need for any assumption to be made about the error term distribution. Instead, DFA assumes that firms' efficiency is constant over time, while statistical noise tends to balance off over time.

Basically DFA can be derived using two different estimation techniques: fixed-effects model (DFA-F) and random-effects model (DFA-R). The former estimates efficiency employing a dummy variable for each firm. If no time invariant regressors exist, it is not necessary to assume the inefficiency terms are independent from the regressors. However, in the presence of invariant firm attributes, omitted from the model, these will be captured by the fixed effects and

[2] In this context, the moral hazard problem appears whenever the contract between the principal (the regulator) and the agent (the firms) gives the agent incentives to make less effort than is optimal.

therefore confused with the inefficiency term. When invariant regressors exist, it is possible to assume independence between explanatory variables and inefficiency, and to estimate a random-effects model. In this case, the procedure consists in calculating a constant for each firm averaging out (over time) the residuals of the panel estimation. The firm with the smallest average residual is considered the most efficient, and other firms' efficiency measures are computed relative to this yardstick.

In summary, the fixed-effects model does not require the assumption of independence between the inefficiency term and the regressors. The random-effects model, in turn, allows the inclusion of time invariant regressors in the model, although at the cost of assuming that the inefficiency term is independent from the regressors. Both models assume that the inefficiency is constant over time, but this assumption can be relaxed.

III. CONSISTENCY CONDITIONS

A problem faced by regulators willing to apply frontier studies consists in the number of methods available for efficiency measurement of individual firms. The problem is far more serious if the different approaches give mutually inconsistent results. The question then arises: are efficiency studies empirically useful?

In an attempt to answer this question, Bauer et al. (1998) propose a set of consistency conditions which must be met by the efficiency measures obtained from the different methodologies for them to be of some use to regulatory authorities. These measures ought to be consistent in their efficiency levels, rankings and identification of the best and the worst firms, ought to be consistent over time and with the conditions under which the industry evolves, and ought to be consistent with other performance measures employed by the regulators. Specifically, the consistency conditions to be analyzed in this work are:

(i) The efficiency measures generated by the different approaches should have similar means and standard deviations;

(ii) the different approaches should rank firms in a similar order;

(iii) the different approaches should identify, in general, the same firms as the "best" and the "worst"; and

(iv) the efficiency measures should be reasonably consistent with other performance measures.

Other consistency measures mentioned in Bauer et al. (1998) but not considered in this paper are:

(v) individual efficiency measures should be rather stable over time, i.e. should not vary significatively from one year to the other;

(vi) the different measures should be reasonably consistent with the expected results from the industry, given the conditions under which it operates. In the particular case of regulated firms, for example, it is expected that those firms regulated under a price cap mechanism will be more efficient than those regulated under rate-of-return regulation.

Broadly speaking, the first three conditions determine the degree to which the different approaches are mutually consistent, whereas the remaining conditions establish the degree to which the different efficiency measures are consistent with reality. So the last three conditions would be like an "external criterion" for the evaluation of the different approaches.

Results of previous studies

Although there exists a vast literature on efficiency measurement in the utilities sector, few studies try to compare the efficiency measures obtained with the different approaches. Among them are the works of Pollitt (1995), Ray and Murkherjee (1995), and Burns and Weyman-Jones (1996). The first two works compare parametric and non-parametric measures, while the latter compare only parametric measures.

Ray and Murkherjee (1995) apply the DEA methodology to a sample consisting of 123 electricity firms, the same sample used by Greene (1990), though this author applies the stochastic frontier approach (under different assumptions about the distribution of the inefficiency term). From the comparison of both studies it is concluded that DEA results are consistent with those of SPF whenever the inefficiency term has a gamma or half-normal distribution. The consistency result is weaker for other distributions of the inefficiency term.

Burns and Weyman-Jones (1996), in turn, compare the efficiency rankings stemming from DFA-R and SPF methodologies. Correlation between both rankings turned out to be 0.395, rejecting the hypothesis of zero correlation at the 95% significance level. In support of the consistency between the models, the authors show that both approaches identify the same firms as the most or least efficient.

Finally, Pollitt (1995) compares the DEA, DPF and SPF approaches in the case of electric power plants, finding correlations ranging from 0.57 to 0.95. According to the author, the results of the application of the different methodologies reveal a relatively high correlation between the rankings derived from the various techniques, specially a very high correlation between both parametric approaches. However, when the same exercise is performed in the case of base load power plants, lower correlations than in the former case obtains.

The literature is far more extense in sectors other than electricity, and the results are diverse (see Bauer et al., 1998, for a discussion of these results in the financial sector). A detailed analysis of the consistency conditions, however, has not yet been attempted in the utilities sector.

An interesting conclusion reached to by Drake and Weyman-Jones (1996) is that the non-parametric and stochastic approaches provide the lower and upper bounds for the efficiency measures, though unfortunately the range between both bounds is often too large.

In this work, to ensure comparability, the four techniques use the same efficiency concept, the same sample of firms, equal specifications of inputs and outputs, and (for parametric methods) the same functional form.

IV. THE THEORETICAL MODEL

Econometric approach

The theoretical specification of the parametric model is the following:

$$Y = f(L, K, Z)$$

Where Y is output (number of customers), L is labor, K is capital and Z is an i-dimensional vector of exogenous variables which allow for the comparison between firms. The most used form for the production function is the Cobb-Douglas specification (Burns and Weyman-Jones, 1996, employed originally a translogarithmic cost function, though they finally chose a Cobb-Douglas form, for it turned out to be more parsimonious), where the inefficiency term enters the model in a multiplicative way:

$$Y = A\, L^{\beta 1}\, K^{\beta 2}\ \{\Pi_i\, Z_i^{\gamma i}\}\ \exp(\varepsilon)$$

Taking logs on both sides of this equation, the inefficiency term enters the model additively:

$$y = \alpha + \beta_1 L + \beta_2 K + \Sigma_i \gamma_i z_i + \varepsilon$$

where α is $\ln(\alpha)$, β_1, β_2 and γ_i are parameters, y is ln(Y), z_i is $\ln(Z_i)$, and ε is the error term. The systematic part of the model determines the maximum obtainable output with a given set of inputs and environmental variables (Z_i), and is known as the frontier. Conceptually, the production function defines a frontier that envelops the technically feasible levels of output associated with different quantities of inputs and different environmental characteristics[3].

Conceptually, all the parametric econometric approaches calculate efficiency measures based on the residuals of a regression. Namely, one firm's inefficiency is a "residual" concept. One controls over the largest number of variables that is feasible. Some firms use more inputs to achieve the same level of inputs than others, and this is what is meant by inefficiency. Therefore, the accuracy of the obtained measures depends upon the grade up to which it was possible to include in the model the largest number of relevant variables. Given that the residuals not only reflect the relative inefficiency of the firms but also the effects of omitted variables in the model, the lower the R^2 of the regression, the higher the likelihood of some important variable being omitted from the model formulation.

In the next subsection the goal is to describe and discuss the empirical specifications of several models found in the literature, with applications to the electricity distribution sector.

Previous studies

Neuberg (1977), for example, describes four related but distinguishable activities in electricity distribution. In the first place, distribution itself, including equipment maintenance, installations to customers and load dispatch. Secondly, metering and billing. Thirdly, sales activity, including subactivities such as publicity, and, finally, administration.

When specifying the variables on which the electricity distribution firms' costs depend, Neuberg suggests four explanatory variables: number of customers served, total sales in kW/h, kilometers of distribution lines, and squared kilometers of distribution area. Burns and Weyman-Jones (1996) add some explanatory variables more: (i) maximum demand, which determines the total capacity of the system, (ii) customer dispersion in the distribution area, which determines the system configuration, (iii) transformation capacity, which affects the network losses, and (iv)

[3] Following Neuberg (1977), the scale elasticity will be given by the proportional impact of changes in inputs and variables denoting the operational characteristics of the firms (Z_i) on the output level. Economies of scale will be present if $(\Sigma_i\gamma_i + \Sigma_i\beta_i) > 1$.

demand structure, which determines the different capacities with which the lines must operate at different hours.

The conceptual problem to be solved is which of these variables is the output, or whether several of them are to be considered outputs. Neuberg refuses to treat firms in the sector as multiproduct firms, because it is impossible to set a price on the above variables and sell them separately (e.g. once number of customers is adopted as the output, being its price the average annual revenue per customer of the firm, the kW/h cannot be sold separately). Given that the remaining variables cannot be considered outputs (nor inputs for which a price is paid) they can be introduced in the model as specific characteristics of the firms to allow for comparisons among them.

Regrettably the problems do not finish here. In the parametric approaches, the estimation of a Cobb-Douglas- or translogarithmic-type cost function (the most used in the literature) requires information about the prices of all the inputs, including the price of capital. However, this information is very hard to collect. This problem is very common in the literature (see Pollitt, 1995, or Huettner and Landon, 1977; both works apply to the electricity distribution sector), and the usual response is to arbitrarily formulate a cost function without including the price of capital in it. Pollitt estimates the following cost function (the author's notation is followed):

$$\begin{aligned} DAC = \alpha &+ \beta_1 \log SALESC + \beta_2 (\log SALESC)^2 + \beta_3\, MAXRAT + \beta_4 (MAXRAT)^2 + \\ &\beta_5\, CUST + \beta_6\, RESID + \beta_7\, OGKMC + \beta_8\, UGKMC + \beta_9\, TRANSC + \beta_{10}\, WC + \\ &\beta_{11}\, AREA + \beta_{12}\, ODUM + \beta_{13}\, CDUM \end{aligned}$$

where *DAC* is distribution cost in 1000s of US dollars per million kW/h, *SALESC* is sales per customer in million kW/h, *MAXRAT* is the ratio of maximum to average demand, *CUST* is number of customers, *RESID* is share of residential sales in total sales, *OGKMC* is overground distribution circuit km per customer, *UGKMC* is underground distribution circuit km per customer, *TRANSC* is transformer capacity (MVA) per customer, *WC* is wage cost in 1000s of US dollars per employee, *AREA* is service area in squared kilometers, *ODUM* is a dummy variable related to ownership (public=1 or private=0), and *CDUM* is another dummy variable that adopts a value of unity when the firm is from UK, and a value of zero otherwise. Including the last dummy variable, Pollitt seeks an international comparison, though somewhat limited (for he only includes in his sample firm data for two countries: UK and USA[4]), of the productive efficiency.

Huettner and Landon (1977), in turn, estimate the following cost function:

$$\begin{aligned} DAC = \alpha &+ \beta_1 \log TCAP + \beta_2 (\log TCAP)^2 + \beta_3\, UTCAP + \beta_4 (UTCAP)^2 + \beta_5\, NTRANSC + \\ &+ \beta_6\, RESIDC + \beta_7\, COMMC + \beta_8\, INDC + \beta_9\, WC + GDUMs + HDUMs \end{aligned}$$

where *DAC* is distribution cost per kW/h, *TCAP* is total capacity in MW, *UTCAP* is average demand as a ratio of maximum capacity, *NTRANSC* is number of line transformers per customer, *RESIDC* is residential sales per customer in MW/h, *COMMC* is commercial sales per customer in MW/h, *INDC* is industrial sales per customer in MW/h, *WC* is the company wage cost in $/hour, the *GDUMs* are geographical dummy variables, and the *HDUMs* are holding company dummy variables (related to ownership).

[4] To the authors' knowledge, Pollitt's (1995) work is the first to attempt an international comparison in the electricity distribution sector.

In neither of the above two models quality variables have been included among the environmental variables considered, because in general there exist minimum quality standards which the firms must achieve. Should these quality standards not exist, the omission of quality variables in the model might cause some firms to appear with lower cost, not because they are efficient, but because they provide a good or service of inferior quality. In this case, it would be convenient to include in the specification of the model quality variables such as frequency and duration of service interruptions, variations in tension, etc. Quality variables should also be included if, despite the existence of quality standards, these were different across firms. The latter aspect becomes particularly relevant when an international comparison is performed.

Mathematical programming approach

The DEA methodology, introduced by Charnes, Cooper and Rhodes (1978), seeks to determine which units (firms) form an envelopment surface or efficient frontier or empirical production function. The firms that lie on (determine) the surface are considered efficient, whereas the firms below the surface are termed inefficient, and their distance to the frontier provides a measure or their relative (in)efficiency (the proportional reduction of inputs and the proportional increase in outputs that would make the firm efficient can be determined).

There exist basically two types of envelopment surfaces (Ali and Seiford, 1993), the so-called constant returns to scale surface (CRS) and variable returns to scale surface (VRS). Their names indicate that an assumption about the type of returns to scale is associated to the choice of either surface (choice that is implicit in the selection of a particular DEA model, as seen below).

The mathematical programming approach also permits the consideration of the chance of some variables being out of the firm's control: these are the so-called environmental variables. The latter can be particularly relevant in electricity distribution, where many variables can be determined by the regulatory framework, the geography, etc. (See Pollitt, 1995, and Weyman-Jones, 1992).

DEA models can be oriented: (i) to the reduction of inputs –input orientation, or (ii) to the augmentation of outputs –output orientation. Given the peculiarities of the industry under study, this work will only consider input-oriented models. Particularly, two of them will be estimated: he first one, assuming constant returns to scale (DEA-C); and the other, assuming variable returns to scale (DEA-V). Both models will take into account the existence of environmental variables.
The theoretical specification of the DEA-C model consists in an optimization problem subject to constraints, like the following:

$$\begin{aligned}&min\lambda\\&s.to \quad u \leq zU, zX \leq \lambda x, zE = e, z \in R^n_+.\end{aligned}$$

This problem gives as a solution the proportion (λ) in which the observed inputs of the firm being analyzed could be reduced if they were used efficiently. U is a $n*r$ matrix of outputs of the firms in the sample (n denoting the number of firms and r the number of outputs). X is a $n*m$ matrix of inputs of the sample firms (m indexing considered inputs). E is a $n*s$ matrix containing all the information about s environmental variables of the n firms. u, x and e are the observed output, input and environmental variables vectors, respectively, of the firm under evaluation. Finally, z is a vector of intensity parameters (z_1, z_2, ..., z_n) that allows for the convex combination of the observed inputs and outputs (in order to build the envelopment surface).

To obtain the second model, DEA-V, it suffices to add the following constraint to the above problem (Seiford and Thrall, 1990):

$$\sum_{i=1}^{n} z_i = 1.$$

In the next subsection, some previous studies are reviewed which shed light over the choice of inputs, outputs and environmental variables to be included in the analysis.

Previous studies

In one of his works, Weyman-Jones (1992) [W-J] uses a DEA model to measure the technical efficiency of a sample of 12 electric utilities in the United Kingdom in the period 1970-1:1988-9. The purpose of this study is to examine the possibility of calculating relative performance measures for non-competitive firms, thus enhancing yardstick competition. Weyman-Jones presents two different models, which are detailed below:

Model W-J 1

1.1.

Inputs:
1. Number of employees
2. Network size (mains km)
3. Transformer capacity (MVA)

Outputs:
1. Domestic sales (kW/h)
2. Commercial sales (kW/h)
3. Industrial sales (kW/h)
4. Maximum demand (kW)

Model W-J 2

1.2.

Inputs:
1. Number of employees

Outputs:
1. Number of customers

Environmental variables:
1. Network size (mains km)
2. Transformer capacity (MVA)
3. Total sales (kW/h)
4. Maximum demand (kW)
5. Population density
6. Industrial share in sales (%)

In the first model, the choice of outputs and inputs follows well-established conventions found in the empirical literature on electricity distribution costs. The second model follows Neuberg's (1977) suggestions. The role of environmental variables is to allow for a productive efficiency measure that explicitly takes into account the differences in the environment under which the firms operate.

In another study, Hjalmarsson and Veiderpass (1992) [H+V] analyze the productive efficiency in Swedish electricity distribution, in the year 1985. The authors estimate the following DEA model:

Model H+V

Inputs:	*Outputs:*
1. Labor (hours)	1. Low voltage output (MW/h)
2. High voltage lines (km)	2. High voltage output (MW/h)
3. Low voltage lines (km)	3. Number of low voltage customers
4. Transformer capacity (kVA)	4. Number of high voltage customers

[H+V] also estimate a second model in which they eliminate outputs 3 and 4; and a third model in which they eliminate outputs 1 and 2 (keeping 3 and 4). Pollitt (1995), in turn, characterizes the distribution services production as follows:

Model P

Inputs:	*Outputs:*
1. Number of employees	1. Number of customers
2. Transformer capacity (MVA)	2. Residential sales (MW/h)
3. Circuit km	3. Non-residential sales (MW/h)
	4. Service area (sq. km)
	5. Maximum demand (MW)

Inputs 2 and 3 represent the capital factor, and input 1, the labor factor. In another section of his work, Pollitt calculates more efficiency measures, assuming alternatively that input 3 and output 4; the latter plus output 5; the latter plus input 2; and, lastly, every variable except number of customers, are environmental variables.

V. DATA AND ESTIMATION

Data

The raw data used in this work have been obtained from the Secretaría General de la Comisión de Integración Eléctrica Regional (CIER) report, "Datos Estadísticos. Empresas Eléctricas. Año 1994" [CIER, 1996][5]. The database includes information about a large number of variables for a sample of 91 electric companies[6] in South America, corresponding to the year 1994. The sample covers ten countries: Argentina (5 firms), Bolivia (8), Brazil (20), Chile (8), Colombia (13), Ecuador (11), Paraguay (1), Perú (13), Uruguay (1) and Venezuela (11).

With regard to the estimation, and according to the issues discussed in the last section, seven variables have been selected out of CIER's database: total sales[7] (SALES, in MW/h),

---

[5] The lack of some information in the case of Argentina firms was completed with information from CEER's database.

[6] In the case of vertically integrated companies (or with more than one function), the distribution function was separated (as long as it was feasible).

[7] The variable SALES was calculated as Total Sales minus Sales to Other Electric Companies, in order to isolate the distribution activity in the case of integrated firms.

residential customers' share (STRUC, a proxy for market structure), distribution circuit (NETW, in km), transformer capacity (TRANSF, in kVA), service area (AREA, in sq. km), number of customers (CUST), and number of employees (LAB). The estimation also considers the inclusion of GDP per capita as a explanatory variable. Due to the lack of service quality data, these could not be included in the model.

After considering some missing data for several firms and the fact that some of the firms in the sample did not belong to the distribution sector, the final sample used in the estimation reduced to 53 companies, with the following detail: Argentina (2 firms), Bolivia (5), Brazil (14), Chile (4), Colombia (7), Ecuador (8), Paraguay (1), Perú (4), Uruguay (0), Venezuela (8).

Table 1 presents a summary of the final sample, based on the chosen variables.

Table 1
Summary Statistics

Variable	Sample size	Mean	Standard deviation
Total Sales (MW/h)	53	4183764	8704092
Residential customers' share (rate)	53	0.87	0.08
Distribution circuit (km)	53	303628	1131251
Transformer capacity (kVA)	53	1040077	2175250
Service area (km^2)	53	60797	186926
Number of customers	53	624474	1061640
Number of employees	53	911	1717

The initial model to be estimated is similar to the W-J 2 formulation of Section IV. The model is as follows:

Initial Model

Output:
1. Number of customers

Inputs and environmental variables:
1. Number of employees
2. Total sales (MW/h)
3. Market structure (%)
4. Mains km
5. Transformer capacity (kVA)
6. Service area (sq. km)
7. GDP per capita (dollars, 1995)

The differences with Weyman-Jones' (1992) work amount to the use of service area instead of population density and the omission of maximum demand. The variable GDP per capita was included in an attempt to capture the differences between countries. Finally, market structure is calculated as residential customers' share instead of industrial sales' share.

Econometric estimation

When performing international comparisons there exists the inconvenient of having to compare different currencies (and many times, different accounting methods). One possible solution would be to estimate production functions, which require only physical units (this is the path followed by Yunos and Hawdon, 1997). This solution is adopted in the present study.

The first stage in the empirical application consists in trying to identify all the variables that might influence the production function. Determining which variables will finally be part of the model can follow a strategy of "going from the general to the specific". This strategy consists in overparameterizing the model by the inclusion of the largest number of variables that is possible, and then simplifying it in the following manner: sequentially eliminate the least significant variable (whenever it is not significant at a 10% level), and reintroduce in each step the variables eliminated in previous steps to check whether they remain not significant (if they are significant, they should be reincorporated in the model). When this task is performed, a model where every single remaining variable is statistically significant obtains.

This was the strategy followed in the selection of the final model, which will be used for both the parametric and non-parametric estimations. The variables service area, mains km and GDP per capita were discarded. The model thus obtained was:

Final model

Output:
1. Number of customers

Inputs and environmental variables:
1. Number of employees
2. Total sales (MW/h)
3. Market structure (%)
4. Transformer capacity (kVA)

In this model all the variables are expressed in logarithms and are significant at a 10% level. The R^2 of the OLS regression is 0.95 and the estimation is heteroscedasticity robust.

Non-parametric estimation

In order to allow for the comparison of results, the same model as in the last section (*Final model*) was used to perform the non-parametric estimation. As was commented upon in Section IV, two models were estimated (one for each envelopment surface), where one output (number of customers), two inputs (number of employees as the labor input, and transformer capacity as the capital input), and two environmental variables (market structure and total sales) were considered.

The previous variable classification agrees with the usual practice commonly found in the applied literature. Particularly, it follows the broad lines drawn by Neuberg (1977) and taken by Weyman-Jones (1992).

In the results it would be expected to find that the efficiency measures are lower for the DEA-C case than for the DEA-V case (fewer firms are found efficient in the former); and also should be found that firms considered efficient in the first model are also considered efficient in the second. The above two results are confirmed by the results. The efficiency measures mean was .740 and .785 for DEA-C and DEA-V respectively.

A somewhat shocking result that arises in the mathematical programming results found in this paper is the large number of firms that are found to be efficient (22 and 29 in each model, out of 53). It is a known fact that as the ratio of variables to observations diminishes, DEA's capacity to discriminate upon efficient and inefficient firms decreases. In this work, five variables and 53 observations have been considered, which a priori seemed a reasonable ratio. It seems clear that having more firms in the sample, or working with panel data, lower efficiency scores (fewer efficient firms) would be obtained.

VI. CONSISTENCY CONDITIONS ANALYSIS

Four different approaches have been used to estimate the efficiency measures: DEA with constant returns to scale (DEA-C), DEA with variable returns to scale (DEA-V), deterministic parametric approach (DPF), and stochastic parametric approach (SPF)[8]. The consistency conditions sketched in Section III will be now analyzed.

(i) Comparison of the distribution of the efficiency measures across the different approaches

Table 2 presents the main characteristics of the distributions generated by the four methodologies employed.

Table 2

Approach	DEA-C	DEA-V	DPF	SPF
Mean	0.740	0.785	0.502	0.736
Median	0.791	1.000	0.515	0.777
Deviation	0.271	0.274	0.162	0.193
Skewness	-0.405	-0.733	0.566	-0.338
Kurtosis	-1.481	-1.207	0.744	-0.813
Maximum	1.000	1.000	1.000	1.000
Minimum	0.254	0.254	0.179	0.283
Sample	53	53	53	53

As it was expected, average efficiency is higher in the stochastic approach (73.6%) than in the deterministic methodologies (68%). The comparison between the parametric and non-parametric approaches concludes that the latter show a higher mean (76% against 62%), probably reflecting the bias of having two many variables relative to the number of observations.

(ii) Correlation between rankings

Table 3 contains the coefficients of Spearman's ranking correlations, which show the existing relationship between each ranking and the others.

All the correlations are positive and, except for those between DEA-V and the parametric approaches, significatively distinct from zero at a 1% level. According to these results, consistency condition (ii) would be met.

[8] Olson, Schmidt y Waldman (1980) used the Monte Carlo method to examine the relative advantages of different estimation methodologies. Maximum likelihood (ML) turned out to be more efficient than modified OLS (MOLS) when the sample size is greater than 400 (though MOLS was superior in the estimation of slope parameters), whereas MOLS has advantages over ML whenever the sample size is below 200. Given that the sample used in this work has less than 200 observations, MOLS was the chosen method to estimate the production function.

Table 3

Approach	DEA-C	DEA-V	DPF	SPF
DEA-C	1.000	0.735**	0.438**	0.442**
DEA-V		1.000	0.184	0.188
DPF			1.000	0.996**
SPF				1.000

* correlation significatively distinct from zero at a 5% level, two tails.
** correlation significatively distinct from zero at a 1% level, two tails.

To test the null hypothesis that the four populations from where the samples came have identical population medians, the Kruskal-Wallis (non-parametric) test was carried out. This test leads to rejection of the null hypothesis, fact that would be pointing a lack of consistency between he approaches.

To conclude, the evidence is not conclusive whether or not consistency condition (ii) is met.

(iii) Identification of the same firms as the "best" and the "worst"

The upper triangle of the matrix displayed in Table 4 shows, for each pair of approaches, the fraction of firms that both approaches simultaneously classified in the upper quartile (13 firms)[9]. The lower triangle of the matrix shows the same for the case of the lower quartile. It is worth mentioning that if the fraction were purely random, it would be expected to be around 25%.

Table 4

Approach	DEA-C	DEA-V	DPF	SPF
DEA-C		0.828	0.773	0.773
DEA-V	1.000		0.759	0.759
DPF	0.615	0.615		1.000
SPF	0.615	0.615	1.000	

These results appear to imply that condition (iv) is being met. The advantage of knowing if the different approaches are consistent relative to the identification of firms as the "best" or the "worst" is that, even in the case of no fulfillment of the first two consistency conditions, it would be possible to use a mechanism like the one employed by OFWAT (the water sector regulatory

[9] Given that the DEA-C model had 22 firms in the first place of the ranking, and the DEA-V model, 29 firms, the upper quartile could not be strictly considered.

authority in the UK), which publishes the efficiency rankings in the media as a reward or a punishment to the firms.

(iv) Consistency with other performance measures

Partial productivity measures, though theoretically inferior to efficiency frontiers, are used as a complement to frontier analysis. Table 5 shows the performance measures used in checking condition (iv):

Table 5

Measure	Mean	Deviation	Maximum	Minimum	Sample
MWh/ Employee	9866	34818	254591	788	53
Customers/ Employee	1245	1966	14684	359	53

Table 6 displays the correlations between both partial productivity measures and the efficiency measures obtained with the four methodologies employed (the correlation between the two partial productivity measures is .956).

Table 6

Approach	MWh/Employee	Customers/Employee
DPF	0.070	0.260
SPF	0.101	0.297
DEA-C	0.188	0.247
DEA-V	0.156	0.203

The external consistency condition requires the efficiency measures generated by the different approaches to be positively correlated with the partial productivity measures, although the correlations must be far away from unity, for the latter do not control for the presence of environmental variables and other inputs (Bauer et al., 1998). According to Table 6, condition (iv) would be met because all the correlations are positive and far from unity. Moreover, it can be observed that the correlations tend to be higher when customers are used a s output in the construction of partial productivity indices (the same choice as in frontier estimation).

VII. CONCLUSIONS

The present work made an international comparison of the relative efficiency of the firms in the electricity distribution sector. To achieve such a goal, it used different methodologies that allowed the construction of several efficiency rankings, on which a consistency analysis was performed. That analysis showed that, broadly, the different approaches are consistent in their means, rankings and identification of the same firms as the "best" and the "worst" (internal consistency). Moreover, the external consistency condition is met, i.e. there exists a positive correlation between the diverse efficiency measures and the partial productivity indices usually used to measure firms' performances.

Despite the particular results found here on the consistency conditions, the paper underscores the importance of conducting a consistency analysis whenever using efficiency measures in applied regulation.

This kind of job highlights the importance of having homogeneous databases in the different countries in order to make the comparisons. In this sense, it is important to note the work of the Comisión de Integración Eléctrica Regional (CIER), source of the information on which this study was based.

Thinking about the future, it would be fruitful to make a periodical survey of the data as a means of building up a database that allows a continued analysis of the evolution over time of the relative efficiency of the electricity distribution companies.

REFERENCES

Aigner, D., C. Lovell and P. Schmidt. 1977. "Formulation and estimation of stochastic frontier production function models". Journal of Econometrics, Vol. 6.

Ali, A.I. and L. Seiford. 1993. "The mathematical programming approach to efficiency analysis". In Fried, H., Lovell, C. y Schmidt, S. *The Measurement of Productive Efficiency*. Oxford University Press.

Bauer, P., A. Berger, G. Ferrier and D. Humphrey. 1998. "Consistency conditions for regulatory analysis of financial institutions: a comparison of frontier efficiency methods". Journal of Economics and Business, 50.

Burns, P. and T. Weyman-Jones. 1996. "Cost functions and cost efficiency in electricity distribution: a stochastic frontier approach". Bulletin of Economic Research, 48:1.

Charnes, A., W. Cooper and E. Rhodes. 1978. "Measuring the efficiency of decision making units". European Journal of Operational Research, 2 (6).

CIER. 1996. "Datos estadísticos. Empresas eléctricas. Año 1994". Comisión de Integración Eléctrica Regional, Secretaría General, Montevideo, Uruguay, October 1996.

Drake, L. and T. Weyman-Jones. 1996. "Productive and allocative inefficiencies in U.K Building Societies: a comparison of non-parametric and stochastic frontier techniques". Manchester School of Economics and Social Studies, 64, March.

Farrell, M. 1957. "The measurement of productive efficiency". Journal of the Royal Statistical Society, Series A, Part III, Vol. 120.

Greene, W. 1990. "A gamma-distributed stochastic frontier model". Journal of Econometrics, 46.

Hjalmarsson, L. and A. Veiderpass. 1992. "Efficiency and ownership in Swedish electricity retail distribution". Journal of Productivity Analysis, 3.

Huettner, D. and J. Landon. 1977. "Electric utilities: scale economies and diseconomies". Southern Economic Journal, 44.

Meeusen, W. and J. van de Broeck. 1977. "Efficiency estimation from Cobb-Douglas production functions with composed error". International Economic Review, Vol. 18, No. 2, June.

Neuberg, L. 1977. "Two issues in the municipal ownership of electric power distribution systems". Bell Journal of Economics, 8.

Olson, J., P. Schmidt and D. Waldman. 1980. "A Monte Carlo study of estimators of the stochastic frontier production function". Journal of Econometrics, Vol. 13.

Pollitt, M. 1995. "Ownership and performance in electric utilities: the international evidence on privatization and efficiency". Oxford University Press.

Ray, S. and K. Mukherjee. 1995. "Comparing parametric and nonparametric measures of efficiency: a reexamination of the Christensen-Green data". Journal of Quantitative Economics, Vol 11, No. 1, January.

Schmidt, P. and R. Sickles. 1984. "Production frontiers and panel data". Journal of Business & Economic Statistics, October.

Seiford, L. and R. Thrall. 1990. "Recent developments in DEA: the mathematical programming approach to frontier analysis". Journal of Econometrics, 46.

Weyman-Jones, T. 1992. "Problems of yardstick regulation in electricity distribution". In Bishop, Kay and Mayer. *The regulatory challenge.* Oxford University Press.

Yunos, J. and D. Hawdon. 1997. "The efficiency of the national electricity board in Malaysia: an intercountry comparison using DEA". Energy Economics, 1997.

CEER Working Paper Series
To order any of these papers, see instructions at the end of the list. To subscribe to all CEER Working Papers see instructions inside the back cover.

Number	Author(s)	Title	Date (mm/yy)
1	Laffont, Jean Jacques	Translating Principle Into Practice	03/1999
2	Stiglitz, Joseph	Promoting Competition in Telecommunications	03/1999
3	Chisari, Omar Estache, Antonio and Romero, Carlos	Winners and Losers from Utility Privatization in Argentina: Lessons from a General Equilibrium Model	03/1999
4	Rodríguez Pardina, Martín and Martín Rossi	Efficiency Measures and Regulation: An illustration of the Gas Distribution Sector in Argentina	04/1999
5	Rodriguez Pardina, Martín Rossi and Christian Ruzzier	Consistency Conditions: Efficiency Measures for the Electricity Distribution Sector in South American	05/1999

Centro de Estudios Económicos de la Regulación

Solicitud de incorporación a la lista de receptores de publicaciones del CEER

Deseo recibir los ejemplares correspondientes a la serie (marque con una cruz la que corresponda), que se publiquen durante 1999:

a) Working Papers Series (...) impreso (...) e-mail, formato pdf

b) Serie de Textos de Discusión (...) impreso (...) e-mail, formato pdf

Mi nombre es:..

Ocupación:..

Domicilio:..

....

....

Firma

Tenga a bien enviar esta solicitud por correo a:

SECRETARIA CEER
Chile 1142, 1° piso
1098 Buenos Aires
Argentina

O por fax, al 54-11-43797588