Efficiency Measures and Regulation: An Illustration of the Gas Distribution Sector in Argentina

Martín Rodríguez Pardina and Martín Rossi

Working Paper N° 4
ISBN 987-519-036-5
April 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
Telephone: 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´s Authorities:
Lic. Enrique Devoto, First Vicepresident Ente Nacional Regulador de la Electricidad (National Regulatory Agency of Electricity, ENRE)-Dr. Roberto Catalán, President Comisión Nacional de Comunicaciones (National Communications Comission, CNC), Dr. Antonio Estache, Instituto para el Desarrollo Económico del Banco Mundial (Economic Development Institute, World Bank), Dr. César Marzagalli, President Universidad Argentina de la Empresa (UADE), Dr. Omar Chisari, Director “Instituto de Economía” (UADE).
Executive Director: Dr. Martín Rodríguez Pardina
Researchers: Economists Gustavo Ferro and Martín Rossi.
Research Assistant: Mr. Christian Ruzzier.
Secretary: Mrs. Graciela Mattaloni.

CEER Working Paper Series
Efficiency Measures and Regulation: An Illustration of the Gas Distribution Sector in Argentina
Martín Rodríguez Pardina and Martín Rossi
Working Paper N° 4
April 1999
JEL N°: L5, L9

Abstract: The paper develops a cost frontier model of gas distribution and estimates it on data for gas distribution companies of Argentina. Efficiency measures are an important tool for regulators, showing how much a firm can rise its output without using more inputs. In Argentina, the Regulatory Framework of the gas sector establishes that only firms that are efficient can earn a rate of return similar to those activities that bear comparable risk. In this context, the estimate of efficiency measures is an indispensable tool to improve regulation of the privatised utilities.

The first part of the paper consists in a theoretical survey of the existent literature where we discuss recent developments in this field, paying special attention to the case of regulated utilities. Deterministic and stochastic cost frontiers are analysed, both for the cases of cross-section and panel data. In the second part, we use econometrics methods to estimate a cost frontier (deterministic and stochastic) for the gas distribution sector in Argentina. Finally, the efficiency rankings of the companies are estimated.

Affiliation of the authors:
Martín A. RODRÍGUEZ PARDINA
Centro de Estudios Económicos de Regulación (CEER), UADE, Argentina

Martín A. ROSSI
Centro de Estudios Económicos de Regulación (CEER), 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
Telephone: 54-11-43797693
Fax: 54-11-43797588

Efficiency Measures and Regulation: an Illustration of the Gas Distribution Sector in Argentina
Martín Rodríguez Pardina and Martín Rossi

I. Introduction

The Gas Supply Industry (GSI) is one of the most important in all modern economies. This Industry has in the last few years undergone fundamental transformations to its structure technology and ownership. The distribution of natural gas started its activity in 1952 with the creation of the enterprise Gas del Estado (GDE). Since the beginning, GDE had the monopoly in all the steps of the transport, distribution and marketing of natural gas. This enterprise was privatised during 1992, being fractionated into two transport and eight distribution companies.

Actually, the industry is organised with several privately owned companies. These stages present varying degrees of competition with production and supply resting on free entry and competitive forces as the main control mechanisms leading to maximum social welfare. Distribution and transport, on the other hand, remain natural monopolies and, as such, in need of some form of public regulation. The regulation of the monopoly stages of the industry is defined in the regulatory frame Law 24076 under the supervision of ENARGAS.

Modern regulatory regimes are focused on improving technical efficiency through incentive mechanisms. Among these, yardstick competition, is a must. Yardstick competition, originally proposed by Shleifer (1985) requires the horizontal separation of some of the stages of the GSI in order to obtain comparative information on relative efficiency levels. This information can then be used to set up tariffs for the regulated companies allowing some efficiency gains to be passed on to consumers and preserving at the same time incentives for the firms to reduce their own costs.

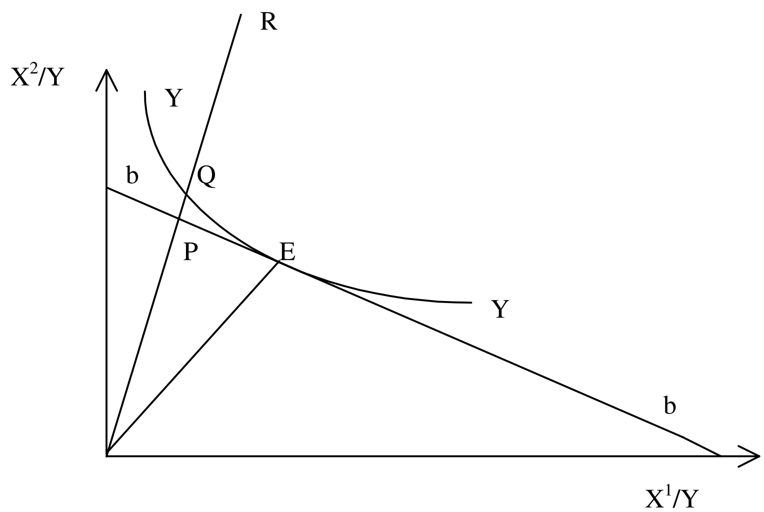
The principles of yardstick competition are quite simple consisting in defining prices, revenues and quality standards based not on data of the own company, as this would eliminate all incentives to improve efficiency, but on the average of a sample of comparable companies. In other words, the regulator acting as the principal prefers to have several agents in order to reduce the existing asymmetry of information. In exchange for this superior knowledge some economies of scale and or scope are lost when the activity is separated into different units.

Efficiency measures are an important tool for regulators, showing how much a firm can rise its output without using more inputs. In Argentina, the Regulatory Framework of the gas sector establishes that firms that are efficient should be able to earn a rate of return similar to those activities that bear comparable risk. In this context, the estimate of efficiency measures is an indispensable tool to improve regulation of the privatised utilities.

II. Efficiency Measures

Efficiency measures were originally introduced by Farrell (1957). Let's assume an industry using 2 inputs X^1 and X^2 to produce a single output Y with a production frontier $Y= f(X^1,X^2)$. This function shows the maximum amount of output that can be obtained by the given set of inputs. If we also assume that $f(X^1,X^2)$ is homogenous of degree one, the technological frontier can be characterised by the unit isoquant (YY in figure 1). Inputs can be measured in output terms X^i/Y. Assuming inputs are bought in competitive markets, the relative price is represented by the slope of the isocost bb, and the firm minimises costs for one unit of output in E, where the marginal technical substitution rate equals the relative price of inputs. By definition, no firm can operate under YY.

FIGURE 1



Let's consider a firm producing at R. This firm is inefficient on two grounds. Firstly, it is operating in a point over the unit isoquant and secondly, it is not using the optimal input combination. Note that the firm in Q has the same input mix that in R using only a fraction OQ/OR of each input (or in other words produces OR/OQ times more output with the same amount of inputs). The ratio OQ/OR is therefore a measure of the technical efficiency of R.

Nevertheless, E and not Q is the optimum production given that although both points represent a 100% technical efficiency, production costs at E are a fraction OP/OQ of the costs at Q (cost of producing at P is the same than in E). The ratio OP/OQ is a measure of allocative efficiency.

Summing up, productive efficiency is the ability of the firm to produce at minimum cost. To achieve minimum costs the firm must use inputs in the most efficient way (technical efficiency) and choose an input mix for which the marginal rate of technical substitution (i.e. the rate at which inputs can be substituted keeping production constant) equals the relative price of inputs (allocative efficiency):

$$\text{Productive efficiency} = \text{Allocative efficiency} \times \text{Technical efficiency}$$

And in terms of figure 1,

$$OP/OR = (OP/OQ) \times (OQ/OR)$$

The measure of productive efficiency adopts values between zero and one with one denoting a firm that is 100% efficient. These measures are defined under the assumption that the production frontier or efficient production function is known. There are basically two possibilities: theoretically defined production function (based on engineering knowledge of the process of the industry) and an empirical function constructed on estimates based on observed data. The usual practice for regulatory purposes is to analyse individual performance in relation with best-observed practice rather than comparing with an ideal practice (which is generally unobtainable). In this work, we will consider that the efficient production function is represented by the best-observed practice among the firms in the sample.

III. Alternative Estimation Methods

Technological frontier studies can be classified according to the specification and estimation methodologies. Focusing on specification, the problem can be viewed from two different approaches: the production function and the cost function. The production function shows the output as a function of inputs, while the cost function shows the total cost of production as a function of output and input's prices.

Gas distribution in Argentina is a regulated activity with companies facing an obligation to supply all demand in their concession area. Since production levels are exogenous to the firm, the objective function for the firms is to minimise costs subject to achieving an exogenously determined product level. Under the prevailing arrangement, the model has to be done with a cost perspective. Indeed, the cost function approach gives consistent estimates when the product is exogenous.

Another advantage of the cost function over the production function approach is the flexibility to adopt different specifications particularly in the cases when the firm produces more than one product. Moreover, estimation of production function allows obtaining a measure of technical efficiency, but ignores allocative efficiency problems. Estimation of cost frontiers, on the other hand, gives information on cost differentials due to technical and allocative inefficiencies. To separate these two effects it is necessary to formulate some additional assumption [(1)].

If we now look at the alternative estimation methods, we see that both production and cost functions estimates can be obtained using statistical or mathematical programming methods. Non-statistical methods estimate frontiers (which can be parametric or non-parametric) without assumptions on the form of the distribution of the error term. The estimates as a result have no statistical properties making it impossible to test hypothesis. In the case of estimates using mathematical programming, the frontier can or not be specified as a parametric function of inputs (obviously, statistical methods are always parametric). The main advantage of non-parametric methods (also known as Data Envelope Analysis or DEA for short) is that no a priori functional form is imposed on the data. The main disadvantage is that to estimate the frontier it uses only a subset of the available data (those actually determining the frontier), while the rest of the observations are ignored.

Once a decision is made on which type of frontier, costs or production, is going to be estimated, and which technique, statistical or mathematical programming, is to be used, the following step is to decided on whether a deterministic or stochastic frontier is to be used. If a deterministic approach is chosen, all observed difference between a particular firm and the frontier is attributed to inefficiency, ignoring the possibility that the performance of a firm might be affected not only by its own efficiency but also by factors beyond its control (such as adverse climate conditions). An additional disadvantage of deterministic estimates is the high sensibility to outliers. A single outlier observation can have strong effects on the results. Moreover, increasing the size of the sample can not solve the problems associated with the “outlier problem”.

The estimation of deterministic frontiers assumes a one-sided error term, implying that it is possible to define exactly the minimum necessary cost to produce a given level of output. Therefore, the actual cost is given by the minimum cost plus an inefficiency term (which by definition is equal or greater than zero). Clearly, the underlying assumption is that all external events which might affect the cost function are the same (and with equal intensity) for all firms.

Starting with the work by Aigner, Lovell and Schmidt (1977) and Meeusen and van de Broeck (1977), the so called stochastic frontier approach is developed, based on the idea that deviations from the frontier can be due to causes not entirely under the control of the firm. This approach uses a mix of one-sided and two sided error terms. Thus, given a level of output, there is a minimum possible cost, which has a random component and therefore can not be exactly determined. The assumption of the models is that the external events affecting the cost function are normally distributed (with the firm facing favourable and unfavourable conditions with a given probability) instead of constant. Once considered this possibility of statistical noise, the remaining is considered inefficiency.

Is important to highlight that one of the advantages of statistical methods over mathematical programming methods is that only the former allows the estimation of stochastic frontiers.

Nevertheless, the estimation of stochastic frontiers is not exempt of problems. Firstly it is necessary to assume a distributional form for the one-sided error term in order to be able to decompose the observed (composed) error term. Secondly, if the deterministic model can label the statistical error as inefficiency, the stochastic approach can have the opposite effect of regarding inefficiencies as statistical noise in those cases in which the sample is free of random errors. The probability that some inefficiencies are erroneously classified as statistical noise is an important problem in the regulatory context: it is at the core of the moral hazard problem for the principals (regulators) in relation to the agents (firms) (Pollit, 1995).

Summarising, parametric mathematical programming and deterministic frontier methods seem to be in the worst world: they are parametric and deterministic. This is why the most used methods in the literature are DEA and stochastic frontier estimates.

IV. Theoretical Model

In traditional cost analysis the problem faced by the firm is to minimise total costs subject to delivering a given level of output. The solution to this problem generates an optimal set of inputs, which depend on output level and input prices. In the same way, it is possible to estimate the cost function of the firm, which depends only on output level and input prices.

The resulting cost model specification is given by:

$$C = f(Y,Z,P_L,P_K)$$

Where

- C: total cost
- Y: output (number of clients)
- Z: n-dimension vector of exogenous variables
- Pk: price of capital inputs
- Pl: price of labour inputs

The most common specification is the Cobb-Douglas function (Burns and Weyman Jones 1996 originally use a translog cost function but discarded it in favour of a Cobb-Douglas on grounds of being more parsimonious) where the inefficiency terms (ε) enters the model as a multiplicative factor (which turns into additive in the logarithmic form).

$$C = A\, P_l^{\beta 1}\, P_k^{\beta 2}\, Y^{\gamma 0}\, \Pi_i\, Z_i^{\gamma i} \exp^{\varepsilon} \qquad (1)$$

Applying logarithms to both sides we obtain:

$$c = \alpha + \beta_1 p_1 + \beta_k p_{2k} + \gamma_0 y + \Sigma_i \gamma_i z_i + \varepsilon \qquad (1')$$

where β_i and γ_i are parameters to be estimated and small cases represent logarithms of the variables in (1). In this model scale elasticity (η) is given by the proportional impact on costs as a result of a change in output and the variables which denote the operating environment of the firm.

This is so because the scale of the gas distribution company is defined not only by the chosen output (customers) but also by the environmental variables included in the model:

$$\eta = \delta c/\delta y + \sum \delta c/\delta z_i,\ i=1,...,n$$

$\sum \gamma_i < 1$ measures the existence of scale economies.

The systemic part of the model determines the minimum obtainable cost with a given set of outputs and environmental variables and is known as the cost frontier. According to the deterministic approach, the stochastic part of the model is completely include in the (in)efficiency term. Given that actual costs can not, by definition, be lower than the frontier cost, the error term can not be negative. Conceptually, the cost function defines a frontier, which envelops the technically feasible costs associated to particular sets of inputs and environmental characteristics.

Estimating these models with (OLS) provides a consistent estimate of the β and γ parameters although the constant term will be biased and inconsistent given that the mean of ε is not zero (Hunt-McCool and Warren 1993). A possible strategy to solve this problem is to estimate the slope parameters using OLS and then correct the constant term in such way as to ensure that all residuals are non-negative and at least one equals zero. The transformed constant term is then a+min(ε_i) (note that min(ε_i) is negative) where a is the OLS estimate of α. This strategy is known as corrected ordinary least squares (COLS) and the i-th firm's cost efficiency can be measured by:

$$\xi_i = \varepsilon_i - \text{Min}\ (\varepsilon_i)$$

and

$$\text{Firm's i efficiency} = \exp(-\xi_i)$$

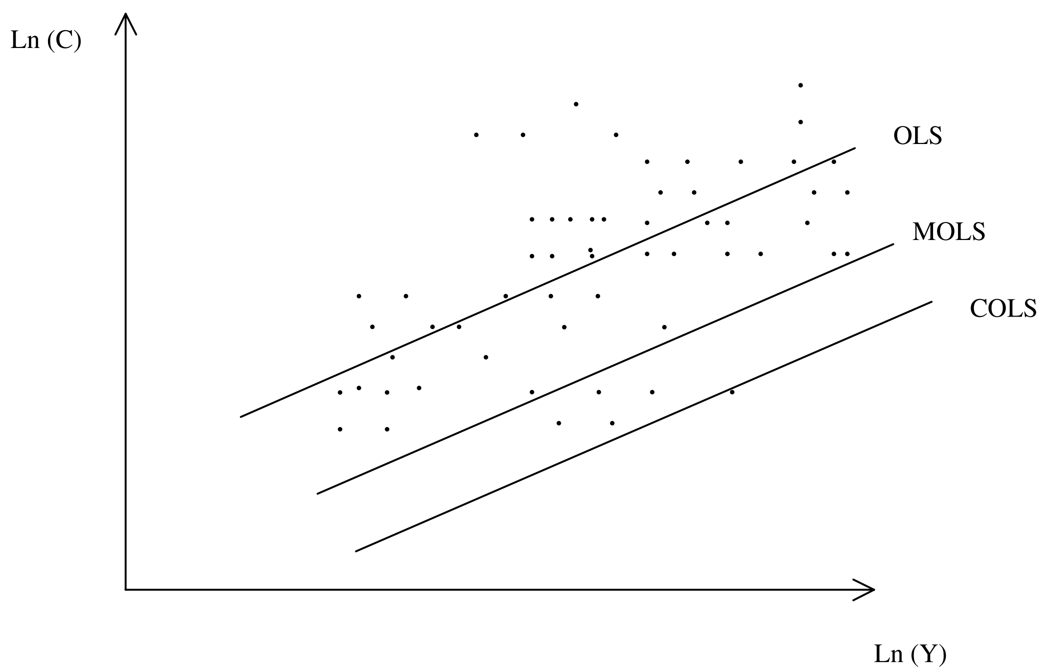
The firm with the min(ε_i) will be 100% efficient. For this firm ξ_i is zero and therefore exp(ξ_i) equals one. The larger the inefficiency of a particular firm i the term ξ_i will be larger and the resulting efficiency measure closer to zero.

This approach has the advantage that there is no need to assume any particular behaviour for the error term. On the other hand, it has the problem (common to all deterministic frontiers) of attributing all deviations from the frontier to inefficiency. Moreover, given that α is changed keeping the OLS estimated β unaltered, the technological structure of all firms (regardless of their efficiency level) is assumed to be the same.

Following Lovell (1992), two alternative strategies for estimating stochastic frontiers can be used: Modified Ordinary Least Squares (MOLS) and Maximum Likelihood (ML). Estimates using MOLS require the formulation of an assumption regarding the distribution of the error term. Several distribution forms have been proposed such as half-normal, exponential, beta and gamma. The procedure has two steps. First the slope parameters are estimated with OLS and then the constant is modified displacing it in a magnitude equal to the average ε_i, which is calculated using the moments of the residuals of the OLS. The OLS residuals are modified in the opposite direction and used to calculate the efficiency measure of each one of the firms in the sample. This strategy does not guaranty that the efficiency measure is in the 0-1 range [(2)].

It is worth highlighting that if the aim is to establish an efficiency ranking of the firms in the sample the results of using MOLS or COLS will be the same. OLS, COLS and MOLS are illustrated in figure 2. A third possible strategy is to use ML estimates. The OLS estimates of the β and γ will, in general, be less efficient than the ML estimators because the latter embodies the a priori information on the asymmetry in the distribution of ε. The efficiency gains from using ML estimates instead of OLS are a function of the skewness of the distribution of the error term, a strictly empirical problem. The same than in MOLS, to estimate using ML it is necessary to specify the distribution of the error term.

FIGURE 2



In the case of stochastic frontiers, the cost function is similar to the one presented in (1') only that now the error term ε is no longer equal to inefficiency but is decomposed into two terms:

$$\varepsilon_i = u_i + v_i$$

where $u_i > 0$ and v_i is not restricted. The v_i captures the effects of statistical noise and are assumed to be independently and identically distributed with an $N(0,\sigma^2_v)$. The u_i error term represents cost inefficiency and is assumed to be distributed independently of the v_i and the regressors. The same than in the case of deterministic frontiers, several functional forms have been proposed for the inefficiency term: half-normal (Aigner, Lovell and Schmidt, 1977), truncated normal (Stevenson, 1980), Gamma (Green, 1990) and exponential (Meeusen y van den Broeck, 1977). The most common distribution used in empirical tests is the half-normal.

Estimation of stochastic frontiers can be done using MOLS or ML, but not COLS. The MOLS method requires an estimate of (1) using OLS, which renders consistent estimates of the slope parameters. The efficiency component can not be directly observed but it can be inferred from the composed error term ε_i. Jondrow et. al. (1982) presents an explicit method to decompose the error term when u_i has a half-normal distribution. The expected value and the distribution mode of the inefficiency term conditioned by the compound error term can be both used as estimators of u_i:

$E(u_i/\varepsilon_i) = \sigma\lambda/(1+\lambda^2)\{\varphi(\varepsilon_i\lambda/\sigma)/\Phi(-\varepsilon_i\lambda/\sigma) - \varepsilon_i\lambda/\sigma\}$,

$M\ (u_i/\varepsilon_i) = \varepsilon_i\ (\sigma^2_u/\sigma^2)$, if $\varepsilon_i \geq 0$

$M\ (u_i/\varepsilon_i) = 0$, if $\varepsilon_i < 0$

where $\sigma=(\sigma^2_v+\sigma^2_u)^{1/2}$, $\lambda=\sigma_v/\sigma_u$, φ and Φ are the probability density function and the cumulative density function of the standard normal distribution, respectively. It is worth highlighting that the distribution of the ε term is asymmetric and non-normal, with the asymmetry characterised by the parameter λ. The greater the values of λ more pronounce the asymmetry. The parameters σ_v and σ_u can be calculated form the OLS moments (see Greene, 1993):

$m_2 = \sigma^2_v + [(\pi/2)-1]\ \sigma^2_u$

$m_3 = -[1-(4/\pi)]\ \sigma^3_u\ (2/\pi)^{1/2}$

Efficiency is given by:

$$\text{Efficiency} = \exp(-u_i)$$

An alternative procedure is based on an estimation of the parameters of the cost function using ML. The error term is then decomposed using the same formula presented above. Olson, Schmidt and Waldman (1980) used a Montecarlo simulation to assess the relative advantages of these two methods. ML results more efficient than MOLS when the sample is larger than 400 (although MOLS is better in the estimation of the slope parameters), while MOLS is more efficient when the sample is under 200. Since in this work the sample is 32 (see chapter V for a description of the data), we use MOLS to estimate the cost frontier.

A problem with the procedure above is that the estimated u_i, although unbiased, are inconsistent. The inconsistency of the estimator of u_i is unfortunate in view of the fact that the main purpose of the exercise is to estimate inefficiency. This problem can be handled using panel data. The principal advantage of panel data is the ability to observe each producer more than once. This allows more flexibility in modelling inefficiency. The simplest model is:

$$c_{i,t} = \alpha + \beta\ X_{i,t} + \varepsilon_{i,t}$$

where $X_{i,t}$ is the independent variables matrix of the model (1') and

$$\varepsilon_{i,t} = u_{i,t} + v_{i,t}$$

There are N firms and T_i observations of each firm (when T is not the same for all the firms we have an unbalanced panel).

If observations on $u_{i,t}$ and $v_{i,t}$ are independent over time as well as across individuals, then the preceding is no different from models of cross section data. The panel nature of the data is irrelevant. But, if one is willing to make further assumptions about the nature of inefficiency, a number of possibilities arise.

If $u_{i,t}$ is assumed time invariant, the model is

$$c_{i,t} = \alpha_i + \beta X_{i,t} + v_{i,t} \quad (2)$$

where $\alpha_i=\alpha+u_i$. This model has the standard form of the models in the panel data literature. In the case regressors are all time-invariant, it is not necessary to assume that the inefficiency term is independent of the regressors. In this case, the model (2) can be estimated like a fixed effects model. No assumption about the behaviour of the inefficiency term is needed. If the most efficient firm in the sample is considered 100% efficient (in a particular moment), $u_i=0$ for that firm, therefore the consistent estimation of α is simply the minimum estimate of α_i. That is,

$$u_i = \alpha_i - \min(\alpha_i)$$

and the cost inefficiency of the i firm is simply $\exp(-u_i)$.

However, in the presence of time invariant attributes of the firms that are omitted from the model, these reappear in the fixed effects, masqueraded as inefficiency (or lack of), when obviously they should be classified otherwise.

In the presence of time invariant regressors, a possibility consists in assuming independence between the independent variables and the inefficiency, and estimate (2) like a random effects model. Given the estimation (random effects) of β, β^*, the individual constants may be estimated from the residuals. If the residuals are defined as $\varepsilon_{i,t}^*= c_{i,t}-\beta^*X_{i,t}$, it is possible to estimate α_i by the mean (over time) of the residuals for the firm i:

$$\alpha_i^* = (1/T)\,(\Sigma\varepsilon_{i,t}^*),\ i = 1, 2, ..., N$$

where the adding up is over t. Then, the estimation is similar to the fixed effects.

A fundamental advantage of panel data is that allows the researcher to test the alternative models (basic, fixed effects and random effects).

If one finds the assumption that inefficiency is time invariant untenable (and it becomes increasingly so as the number of time series observations becomes larger), some structure of how the inefficiency evolves across time could be imposed. One possibility is a two way effects model,

$$u_{i,t} = u_i + r_t$$

This can be treated as a fixed or random effects model, and have the disadvantage that a similar effect hits every firm in each particular period. To drive this inconvenient, Cornwell, Schmidt and Sickles (1990) allow the individual effect to evolve over time, as a quadratic function:

$$u_{i,t} = \gamma_{i,1} + \gamma_{i,2}\, t + \gamma_{i,3}\, t^2 \quad (3)$$

That is, the inefficiency term is a quadratic function of time, but the form is not the same across firms. The estimation is as follows: in the first step one estimates (1') by OLS with estimated residuals $\varepsilon_{i,t}$. In the second step one estimates the γ_i coefficients by estimating the following equation by OLS:

$$\varepsilon_{i,t} = \gamma_{i,1} + \gamma_{i,2}\, t + \gamma_{i,3}\, t^2 + v_{i,t} \quad (4)$$

where $v_{i,t} \sim N(0,\sigma^2_v)$. The fitted values from (4) provide an estimate of $u_{i,t}$ which is the efficiency estimator. Other functions may be used. Also, a model with technological progress can be estimated:

$$c_{i,t} = \alpha + \beta\, X_{i,t} + \omega_t + \varepsilon_{i,t}$$

where technological progress is estimated by the ω coefficient. In the present work T=4, therefore a model like (3) can not be estimated robustly.

V. The empirical estimation

<u>The cost function</u>

In practice, the estimation of cost function is not as simple as the theoretical model implies. The costs of a gas distribution company depend on a variety of factors in addition to output levels and input prices. Neuberg (1977) describes four related but distinguishable activities in electricity distribution that can be assimilated to the gas case. Firstly, distribution properly which includes maintenance of equipment and installations to users and load dispatch. Secondly, meter reading and billing. Thirdly, sales including related activities such as publicity and fourthly administration. Neuberg suggests four variables as main cost drivers in electricity distribution: number of customers served, total KWh sold, Km of distribution lines and Km2 of distribution area. Burns and Weyman Jones (1994) add some additional variables: maximum demand (which determines system configuration and size), transform capacity (which affects losses) and demand structure (which determines load factors at different moments of the day).

The main conceptual problem is to identify within this set of variables which one or ones are the output. Neuberg discards the possibility of treating distribution companies as multiproduct firms given that the different variables can not be separately sold and/or priced. For example, once the number of clients is identified as the product (with a price equal to average annual revenue per customer of the firm), energy sales in (KWh) can not be sold separately. Given that once a product is chosen the rest of the variables can not be treated as outputs, they can be introduced in the model as firm specific characteristics to allow inter firm comparisons.

Unfortunately, the problems do not finish here. The estimation of a Cobb-Douglas cost function requires data on input prices, included the price of capital input. However, this information is very difficult to obtain. This problem is very usual in the literature (see Pollit, 1995 o Huettner and Landon, 1977, both for the electricity distribution sector), and the usual way is the formulation of an arbitrary cost function, without including the price of the capital input. Pollit estimates the following cost function (notation of the author):

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

where DAC is the distribution cost in 1000s of US dollars per million KWh, SALESC is the total sales per customer in million KWh, MAXRAT is the ratio of maximum to average demand, CUST is the number of customers, RESID is the share of residential sales in total sales, OGKMC is the overground distribution circuit Km per customer, UGKMC is the underground distribution circuit Km per customer, TRANSC is the transformer capacity per customer, AREA is the service area in square Km, WC is the wage cost, ODUM is an ownership dummy variable y CDUM is a UK country dummy variable.

Huettner and Landon, on the other hand, estimate the following cost function:

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

where TCAP is total capacity in MW, UTCAP is the average demand as a ratio of maximum capacity, NTANSC is the number of line transformers per customer, RESIDC are the residential sales per customer, COMMC are the commercial sales per customer, INDC are the industrial sales per customer, WC is the company wage cost, GDUMs are geographical dummy variables and HDUM are holding company dummy variables.

The gas distribution sector is similar to the electricity distribution sector and, therefore, the variables to include in the specification of the cost frontier are similar to those used by the preceding authors. The cost of the distribution gas firms is modelled in this paper as a function of pipe kilometres, number of customers, concession area, sales, maximum demand, market structure and the price of the labour input.

The data

The data set is a balanced panel of 8 firms observed over 4 years (1993-96). The set includes information about total cost (COST), kilometres of pipes (KMNET), number of customers (CUSTOM), concession area (AREA), number of employees (EMPL), sales (SALES), market structure (the ratio of residential to total sales, STRUCT), maximum demand (MAXDEM), and the price of the labour input (WAGE). The summary statistics are presented in table 1 and 2.

Table 1
Summary statistics

Variable	Sample	Mean	Standard Error
Total cost ($)	32	122262846	199652230
Km. of net	32	10411	4487
Number of customers	32	625465	520901
Concession area (Km^2)	32	301490	245400
Residential sales (MM m3/day)	32	1.96	1.51
Total gas sold (MM m3/day)	32	8.14	4.61
Maximum demand (thousand of m3)	32	9.59	5.55
Number of employees	32	541	348
Mean wage (annual)	32	32969	4818

The total cost is in pesos and has been obtained from the exhibit H of the balance sheets of the firms [(3)]. The total cot is calculated net of the acquisition of gas. The number of clients, total sales, residential sales and distribution network, were provided by the ENARGAS. Sales are en MM m3/day (9300 Kcal). The annual wage was calculated as the ratio of total expenditure in salaries to number of equivalent employees (the full time employees plus half the part time employees). The maximum demand was obtained as the ratio of the month with maximum demand to the number of days in that month. The concession area was calculated using data provided by INDEC.

Table 2
Correlation matrix

	Cost	Wage	Sales	Maxdem	Kmnet	Estruct	Custom	Area
Cost	1	-0.0096	0.7438	0.8195	0.3045	0.3287	0.9305	-0.5384
Wage		1	0.0959	0.0600	-0.1826	-0.0168	-0.0617	0.0645
Sales			1	0.9434	0.6139	0.2713	0.7861	-0.1361
Maxdem				1	0.6088	0.4139	0.8683	-0.2736
Kmnet					1	0.5289	0.6040	-0.2500
Estruct						1	0.4650	-0.0877
Custom							1	-0.6362
Area								1

<u>The estimation</u>

The initial function to be estimated is in line with the works of Huettner and Landon (1977) and Pollit (1995), although the estimated costs are total rather than average:

Model 1:

Log COST = α + β log WAGE + γ_0 log CUSTOM + γ_1 log KMNET + γ_2 log AREA + γ_3 log SALES + γ_4 log STRUCT + γ_5 log MAXDEM

In table 3 are presented the results of the OLS estimation.

Table 3
Dependent variable: Total cost

Independents variables	Model 1 Coefficients (t statistics)	Model 2 Coefficients (t statistics)
Constant	3.854	0.008
	(0.73)	(0.00)
Ln(WAGE)	0.035	0.135
	(0.16)	(0.75)
Ln(CUSTOM)	2.010	2.272
	(5.63)	(10.57)
Ln(KMNET)	-1.514	-1.664
	(-6.08)	(-8.89)
Ln(AREA)	0.094	0.158
	(1.11)	(3.41)
Ln(SALES)	0.296	
	(0.92)	
Ln(ESTRUCT)	0.351	0.317
	(4.81)	(5.07)
Ln(MAXDEM)	0.067	0.310
	(0.24)	(3.52)
Estimation method	OLS	OLS
Number of firms	8	8
Number of years	4	4
Sample	32	32
R^2	0.98	0.98
F statistics	309.72	363.42
Prob. (F stat.)	0.0000	0.0000

White test accepts the null hypothesis of no heterocedasticity.

The coefficient of SALES is not significant at the usual levels. Moreover, since this variable is correlated with MAXDEM (0.94), CUSTOM (0.78) and KMNET (0.61). To avoid problems of multicolineality, this variable was dropped and model 2 was estimated:

Model 2:

Log COST = $\alpha + \beta$ log WAGE + γ_0 log CUSTOM + γ_1 log KMNET + γ_2 log AREA + γ_3 log STRUCT + γ_4 log MAXDEM

Efficiency figures are calculated on basis of model 2. The slope parameters are all significant at the usual levels of confidence. Only the constant and WAGE are not significant, although this last one has the right sign (a raise in wages produces an increment in total cost).

The individual annual figures were calculated subtracting the minor residual of the year to the OLS residuals. The annual figures obtained were averaged to obtain the final score. The results are presented in table 4.

Table 4
Firm's efficiency: deterministic frontier
1993-1996

Firm	1993	1994	1995	1996	Mean (ranking)
1	0.838	0.808	0.765	0.793	0.801 (7)
2	0.822	0.786	0.884	0.952	0.861 (3)
3	1.000	0.776	0.710	0.945	0.858 (4)
4	0.929	1.000	1.000	1.000	0.982 (1)
5	0.853	0.788	0.771	0.844	0.814 (6)
6	0.851	0.7779	0.812	0.880	0.830 (5)
7	0.838	0.872	0.873	0.872	0.864 (2)
8	0.658	0.716	0.737	0.869	0.745 (8)

An assumption under the above proceeding is that there were no structural changes in the technological parameters during the analysed period.

Using the procedure presented in section IV, the measures of efficiency corresponding to the stochastic frontier were obtained (assuming that the inefficiency term follows a half normal distribution). The calculation was done with the conditional mode, rather than with the conditional mean.

Given that the skewness of the OLS residuals was negative (i.e., the wrong sign), both the mean and the variance of the inefficiency term are equal to zero (all the firms are operating on their frontier, i.e., are 100% efficient), which could be showing that the data is inconsistent with the specified stochastic frontier (Waldman, 1982).

The final step consists in the estimation with panel data techniques. The model estimated is similar to model 2, but without the variable AREA (the OLS estimation of this model is presented in table 5). The idea is that this variable can be removed from the model because it is only a proxy of the service area (there are firms that service a small area but have a large concession area).

Table 5
Dependent variable: Total cost

Independents variables	Basic model Coefficients (t statistics)	Fix effects model Coefficients (t statistics)
Constant	6.833	
	(2.37)	
Ln(WAGE)	0.023	0.513
	(0.11)	(4.13)
Ln(CUSTOM)	1.577	-0.434
	(19.3)	(-0.58)
Ln(KMNET)	-1.128	0.771
	(-9.34)	(1.47)
Ln(ESTRUCT)	0.435	0.077
	(4.57)	(0.54)
Ln(MAXDEM)	0.246	0.144
	(3.52)	(1.19)
Estimation method	OLS	OLS
Number of firms	8	8
Number of years	4	4
Sample	32	32
R^2	0.98	0.99
F statistics	307.90	3392.40
Prob. (F stat.)	0.0000	0.0000

The panel data specification were test as usual, and the fixed effects model was preferred (see table 5). The assumption behind this model is that the inefficiency is time invariant (this could be reasonable, a priori, considering the length of the analysed period, though also has to be considered that the sector is in the post privatisation period and, therefore, subject to important changes) and, moreover, that it is the only invariant attribute. The latter assumption is extremely dubious (the firms have attributes such as capital stock, location, or some other characteristics that not vary over time). Another important assumption is that, as the sample size increases, it must become more likely that firms on the estimated frontier are near the true frontier. Having this in mind, table 6 presents the efficiency figures associated with the fixed effects model.

As can be observed, and probably related to the previous argument, the efficiency ranking is very different from the measures presented in table 4.

Table 6
Measures of efficiency: fixed effects model
1993-1996

Firm	Efficiency (ranking)
1	0.039 (8)
2	0.348 (7)

3	0.559 (6)
4	0.735 (5)
5	1.000 (1)
6	0.979 (3)
7	0.997 (2)
8	0.957 (4)

VI. Conclusions

In Argentina, the Regulatory Framework of the gas sector establishes that firms that are efficient should be able to earn a rate of return similar to those activities that bear comparable risk. In this context, the estimate of efficiency measures is an indispensable tool to improve regulation of the privatised utilities.

In this paper the efficiency average ranking per firm has been estimated for 3 alternative models: stochastic frontier, deterministic frontier and fixed effects model. The differences in the rankings can be showing the limitations of this model to explain the behaviour of the sector in the analysed period. The time invariant assumption of the efficiency term between the years 1993-96 is questionable because the special characteristics of this period (post-privatisation). It may be useful to relax the assumption of fixed inefficiencies over time and allow the individual effect of each firm to evolve over time. This is impossible with the available data and would be important to have this in mind for further research.

Notes

(1) The regulatory framework of the gas sector in Argentina makes reference to efficiency (as a whole) rendering the estimation of cost functions more useful than that of production functions.

(2) If an observation has a negative and sufficiently large OLS residual, an efficiency measure larger than 1 is possible. In this case the theoretical concept of cost frontier as the minimum attainable value is not fulfilled.

(3) Costs were taken at current values (not adjusted for inflation). The short span of the sample (4 years) and the low inflation during the period made it unnecessary to use constant values.

References

Aigner, D., Lovell, C. y Schmidt, P. (1977): *Formulation and estimation of stochastic frontier production function models*, Journal of Econometrics, Vol. 6.

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

Cornwell, C., Schmidt, P., y Sickles R. (1990): *Production frontiers with cross-sectional and time series variation in efficiency levels*, Journal of Econometrics, Vol. 46.

Farrell, M. (1957): *The Measurement of productive efficiency*, Journal of the Royal Statistical Society, Series A, Part III, Vol. 120.

Green, W. (1993): *The econometric approach to efficiency analysis*, en The Measurement of Productive Efficiency, Fried, Knox y Schmidt (editors), Oxford University Press, Oxford.

Huettner, D. y Landon, J. (1977): *Electric utilities: scale economies and diseconomies*, Southern Economic Journal, 44.

Hunt-McCool, J. y Warren R. (1993): *Earnings frontiers and labour market efficiency*, en The Measurement of Productive Efficiency, Fried, Knox y Schmidt (editors), Oxford University Press, Oxford.

Lovell, C. (1993): *Production frontiers and productive efficiency*, en The Measurement of Productive Efficiency, Fried, Knox y Schmidt (editors), Oxford University Press, Oxford.

Meeusen, W. y van de Broeck, J. (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., Schmidt, P. y Waldman, D. (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 privatisation and efficiency*, Oxford University Press.

Shleifer, A. (1985): *A theory of yardstick competition*, Rand Journal of Economics, Vol. 16, 3, Autumn.

Stevenson, R. (1980): *Likelihood functions for generalised stochastic frontier estimation*, Journal of Econometrics, Vol. 13.

Waldman, D. (1982): *A stationary point for the stochastic frontier likelihood*, Journal of Econometrics, 28.

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

Centro de Estudios Económicos de la Regulación

1.
2.

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