SECTOR FRAGMENTATION AND AGGREGATION OF SERVICE PROVISION IN THE WATER INDUSTRY[1]

FORTHCOMING, JOURNAL OF PRODUCTIVITY ANALYSIS

Maria Luisa Corton

Visiting Assistant Professor of Economics

University of South Florida at Saint Petersburg

Research Associate

Public Utility Research Center

University of Florida

Phone: (352) 214-5132

mcorton @ ufl.edu

Abstract

This study analyzes the structure of the water industry in Peru. It investigates economies of scale and cost inefficiencies in 43 water providers during the years from 1996 to 2005. Different water sources in different geographical regions directly affect the production technology of the sector, making a regional cost frontier model appropriate. In 2005, 48.7% of the population of Peru lived in poverty, and 43.87% of the water produced was lost. These large proportions support the assumption that water producers allow water to be lost as a way to satisfy the water demands of the poorest segment of the population. Findings indicate a cost rise of 0.10% for each 1% increase of joint production of both outputs: water lost and produced. This result may be suggestive of a price the utility pays for allowing unauthorized connections in the network. From the perspective of the board of directors, this cost increase is the political cost of gaining municipal votes from the poorest segment of the population. Economies of scale are larger for utilities operating in the forest than in other regions. This has important implications for any future consolidation process that may be undertaken when the sector is opened to private participation.

Keywords: Panel data analysis, stochastic frontier, cost efficiency, scale economies

JEL Classifications: C23, C51, L11, l51, L95

---

[1] I am indebted to Sanford Berg for his advice and suggestions at various stages of this research. I am grateful to David Sappington for his comments on critical aspects of earlier drafts. I am grateful to two anonymous referees for their valuable comments and suggestions. I would like to acknowledge the cooperation of Ernesto Bazan Asencios and Juan Francisco Cordova during the data review process (both at SUNASS). Thanks to Patricia Casey for her editorial assistance. The usual disclaimer applies.

1. INTRODUCTION

In 1995 the government of Peru restructured the water industry by decentralizing the provision of water service to municipalities. The purpose was to expand the service provision as well as to stimulate efficiency among service providers. The water utilities, previously owned by central government, are now owned by local municipal authorities. In addition to the change in ownership, the decentralization process implied more responsibilities for the water regulatory agency, SUNASS (Super Intendencia Nacional de Servicios de Saneamiento). The previous cost-plus type regulatory contract was replaced by an incentive contract, making the regulatory tasks more complex. After decentralization SUNASS collects annual data from the water utilities by means of "master plans" with the objective of measuring their efficiency.[2] In addition, the regulatory rules in place give SUNASS the power to merge or split utilities, if necessary, to create a more efficient market structure.

By early 2000 the government had exhausted its capacity to fund infrastructure projects, and initiated a series of actions to attract private capital into the water sector. A concession bid procedure started in 2005 targeting one of the largest utilities in the sector. The motivation for the current analysis comes from the concerns expressed by SUNASS in a report assessing the development and performance of the sector by the end of 2006 (SUNASS 2006). In this report, SUNASS identifies several aspects critical for the water sector such as lack of investment and fragmentation, which means the presence of a large number of small service providers in the sector.[3] Beyond the financial sustainability of these small utilities, the concerns of SUNASS were related to their fundraising capability, and their limited attraction to private investors.

The objective of this study is to analyze the structure of the water industry in Peru by investigating the presence of economies of scale and cost inefficiencies of the water utilities while considering the production particularities of the sector. As a natural monopoly, economies of scale should not be a surprise, but the extent of these economies would be of interest to policy makers as they assess the optimal market structure and open the sector to private participation.

---

[2] See Tamayo, Barrentes, Conterno and Bustamante (1999) and Corton (2003) for detailed descriptions of the industry situation in 1995, and the setting up of the benchmarking system. The master plans are comprised of accounting, operating, and commercial data as well as future network expansion and maintenance.

[3] From the 49 service companies comprising the sector in 2006, approximately 39% provided service to less than 10,000 connections, while only 20% provided service to more than 40,000 connections.

A crucial characteristic of the water industry in Peru is that service providers operate in three distinct geographical regions: the *mountains*, the *forest* and the *coast*. The source of water in these regions may be surface or underground or a mix of both, which has a direct effect on the production technology of this industry. Consequently, the economic model is specified as a regional cost function with input prices and output coefficients representing cost deviations from a base region.

Another particularity of the Peru water industry is the large volume of water lost: 43.87% in 2005. For water utilities in developed countries this value average 15% and it is 35% for utilities in developing countries (Kingdom, Liemberger and Marin 2006). For the same year, the World Bank reports that 48.7% of the population in Peru lived in poverty.[4] Assuming that water lost originates from water connections non-authorized in the network system, controlling water lost entails denying the service to the poorest and largest segment of the population. The board of directors of water utilities may find that satisfying the water needs of this segment is advantageous, hoping to secure votes during municipal elections. In the framework of this analysis, unwillingness to deny such service explains the large proportion of water lost. Thus, in this study water lost is considered as an output together with water produced.[5]

The study utilizes a panel data of 43 water service companies for the period 1996 to 2005. The source of the data is the set of master plans reported by these utilities to SUNASS. The analysis considers a long-run scenario based on the estimation of a stochastic cost frontier in which inefficiency is allowed to vary over time. Results from the analysis show the production technology to be homothetic, and not homogenous implying that the input mix is constant with scale and that returns to scale vary with the scale of the firm and the production input mix.

Among other results, the study supports the joint production of water produced and lost given the positive and statistically significant coefficient for the interacted water lost and produced term. A 1% increase of joint production implies an increase in costs of 0.10%. It is plausible to interpret this cost increase as the utilities' cost for allowing non-authorized

---

[4] Source: http://go.worldbank.org/AHUP42HWR0 (last visit 02/26/2010). The World Health Organization-Statistics show that by 2007, Peru had a population of 27.9 million people with 71% living in urban areas. During 2000 to 2007, access to drinking water in urban areas increased by only1%, while in rural areas the increase was 7%. On average, water utilities deliver service to 82% of the total population in their area of service. http://www.who.int/whosis/whostat/2009/en/index.html

[5] Two leading studies on this issue are Garcia and Thomas (2001), who found a tradeoff between water produced and network leaks in the French water industry, and Antonioli and Filippini (2001) who included percentage of water loss as a firm specific characteristic in the analysis of the Italian water industry.

connections in the network. For the board of directors of the utility, this result might reflect the cost of capturing municipal votes from the poorest segment of the population. Finally, this cost increase may be interpreted as the cost of exerting stricter network control and maintenance.

The study is organized as follows. Section two summarizes industry framework. Section three describes the model specification and the estimation procedure. Section four presents the empirical results. Section five concludes.

2. INDUSTRY FRAMEWORK

The description of the production process of this industry follows the five steps that Garcia and Thomas (2001) utilize to characterize the French water industry. The first step, *production and treatment*, comprises water extraction, if underground, and water preliminary treatment such as disinfection, filtering, and different degrees of softening for both surface and underground water. Water as a commodity has no cost in Peru. The second step is the *transportation of water* from the production facilities to the distribution city gate through transmission pipelines. *Storing water* in facilities such as water tanks or towers is the third step, and the fourth step is *pressurization of water* from the storage facilities into pipelines. The final step is the *distribution* of water from the city gate to customers. Water utilities in Peru are vertically integrated, as they perform each of these production steps. [6]

A particularity of the water industry in Peru is that each service provider is located in one of three natural geographical regions of the country. These regions are classified according to vegetation and climate as *mountains, forest,* and *coast.*[7] The water source in each region varies from surface to underground. In general, surface water does not need pumping but needs a more intense water treatment to meet water quality standards. When the source of water is in the mountains, gravity promotes water flow, while the required amount of pumping is higher when serving customers in a flatter geography like the coast. These differences have a direct impact on the utilities' production technology and subsequently, on the utilities' operating costs.

---

[6] Yet, the accounting books of these utilities do not reflect the costs from each step separately. However, from an accounting perspective, it is plausible to assume that the costs related to the production process are spread among the accounts: cost of sales, sales expenses and administrative expenses.

[7] From a total of 43 utilities considered in the analysis, 16 are located in the mountains, 9 in the forest, and 18 on the coast. The coast is the most populated region with a population growth rate of 18% for the period 1996 to 2005, and the mountain region is the less populated with only 9% growth rate for the same period.

Another characteristic of the water industry in Peru is the size of the utilities. SUNASS classifies water utilities as *large*, *medium*, and *small* depending on the number of water connections served. [8] All but one large utility are located on the coast, and the mountains are served only by small and medium size utilities. The forest is served by all three sizes of utilities.

Regarding the outputs and inputs in this industry, the most frequently identified outputs in the empirical literature are volume of water sold (generally classified by type of customer), population served, number of connections, and volume of water produced.[9] For the identification of outputs this analysis follows Neuberg (1977) in his argument about the "separate marketability of components" as a necessary (but not sufficient) condition to identify an output.[10] According to this classification, two outputs are identified in the water industry of Peru: volume of water produced and volume of water lost. [11] From 1996 to 2005, volume of water was the only product billed to customers. [12] Water lost, on the other hand, is considered an output because it is assumed to be valued by a large segment of the population relying on non-authorized connections to satisfy their water needs. It is plausible to assume that controlling water lost entails denying

---

[8] Large companies are those with more than 40,000 connections; medium size companies are those with connections between 10,000 and 40,000; small companies are those with less than 10,000 and more than 1,000 connections. Of the 43 utilities, 8 are large, 19 are medium, and 16 are small. Between 1996 and 2005, only two utilities changed their size. For the purpose of this study, utilities are classified according to the number of water connections registered by SUNASS in 1996.

[9] Aubert and Reynaud (2005) utilize volume of water delivered and the number of customers served in a variable cost function for the Wisconsin water sector. Saal and Parker (2004) consider water volume delivered and the number of connections in a study of the England and Wales water industry. Garcia and Thomas (2001) define volume of water lost and volume of water delivered in the French water sector study. Saal and Parker (2000) utilize water supply and population served in their study of the England and Wales water sector. Renzetti (1999) identifies water sold to residential and non-residential customers in the analysis of cost supply and pricing practices of the water and sewerage utilities in Ontario.

[10] Some of the authors using this definition in the electricity sector are as follows: Hattori, Jamasb and Pollit (2005); Lowry, Getachew and Hovde (2005); Jamasb and Pollit (2003); Hattori (2002) and Rossi (2001).

[11] Given that all utilities provide water and sewerage services, a previous version of this paper included number of sewerage connections as an output. However, multicolinearity was detected and the number of variables related to the service was very restricted. Additionally, sewerage was never billed during the analyzed period. All in all, the sewerage sector is not represented in the analysis.

[12] All customers in this study are assumed to be residential given the unavailability of data for type of customers. Although the number of water connections is an output in several water studies, including number of water connections together with volume of water introduces multicolinearity in the model. A large correlation coefficient for the variables, 0.9893, implies that both represent the same type of information as they are strongly linearly related.

service to a relatively large share of the population given the poverty conditions of the population during the analyzed period. The board of directors of these utilities might find it is advantageous to satisfy the water needs of this large and poorer segment of the population, and to secure their votes during municipal elections. Volume of water lost is defined as the difference between volume of water produced and billed.

Regarding input factors of production, labor and capital are included in the analysis with labor comprised of two types of workers: direct and indirect.[13] On average, labor costs represent 40% of utilities' operating costs. Labor is classified according to the contractual obligations acquired by the utility at the time of hiring. *Direct labor* is comprised of workers who have a permanent position in the company and are entitled to the labor benefits of the company. *Indirect labor* is comprised of workers under explicitly and limited terms of employment regarding salary. This group has no working benefits other than a monetary amount assigned at the outset of the work agreement. As in other network industries, network length[14] is used as a proxy for capital. Overall, the rationale for using network length in the water industry is that the amount of capital necessary to lay down pipes is significantly higher when compared to the capital needs for other types of network developments such as installing pumps or treatment facilities. Data for these capital items were not available.

3. MODEL SPECIFICATION AND ESTIMATION

3.1 Economic Model

This analysis assumes that the managers of these utilities act to minimize long-run costs. This implies that all input factors are free to adjust and output is exogenously determined by the regulatory obligation to supply customers. This assumption may seem strong in the context of this industry. However, municipal leaders stay in power only for one year and the analyzed period is ten years, so it is assumed that any differences among managerial objectives, other than cost minimization, are smoothed out without distorting the free adjustability of input factors. In

[13] Non-labor data such as materials or energy consumption are not available.

[14] The length of the network includes transmission and distribution pipes up to the customer connection measured in kilometers. The measure is performed from the initial point of the transmission network, characterized by bigger pipes, up to the end of the larger segment of the distribution network.

addition, the period depicts a substantial expansion of the networks, suggesting enough flexibility of capital, and supporting a long-run analysis. [15]

In addition, a long-run analysis requires the use of price of capital instead of capital stock. The companies in Peru are not an exception regarding the difficulties Latin American companies face when calculating the price of capital. For this reason, in this analysis a proxy for the price of capital is defined as finance expenses plus depreciation divided by length of network.[16] The rationale for using this proxy is that an increase in debt will occur when the utility expands service, which implies an increase in length of the network. Finance expenses will then reflect this expansion of service. Thus, this proxy captures characteristics related to debt financing managerial capabilities, and the possible value of investing in network developments. Nevertheless, this proxy is under-representing the relative risk of investing in this sector as opposed to investing in other industries. Prices for direct and indirect labor are calculated as annual labor costs divided by the number of workers. It is assumed that on average, the number of hours worked by indirect and direct workers is similar.

Costs in this analysis comprise sales costs, sales expenses, administrative expenses, annual depreciation, and finance expenses, which represent total operating annual costs. The source of the cost figures is the accounting books of the utilities, and it is assumed that the accounting definitions adopted by all firms are the same. For the period 1996 to 2005, large utilities have a mean value for total operating annual costs of $41,197.[17] For medium and small companies, these values are $2,455 and $569, respectively.

Given multiple outputs and price of inputs, a transcendental logarithmic functional form is appropriate to represent the technology.[18] This flexible functional form does not place a priori restrictions on substitution among factors of production, and it allows scale economies to vary with level of output, not imposing homotheticity or unitary elasticity of substitution. These

---

[15] Between 1996 and 2005, large companies show a network expansion of 30.74%, medium companies expanded by 56.69%, and small companies by 50.82%.

[16] Given the regulatory environment, it is assumed that the depreciation procedure is similar and based on similar estimation of asset values.

[17] These figures are expressed using current US dollars. Exchange rates for the period were obtained from Peru's Central Bank web site: http://www.bcrp.gob.pe (last visit 02/20/2010).

[18] This functional form, as opposed to the Cobb-Douglas functional form, avoids cost curvature issues when multiple outputs are employed. At a 0.001level of statistical significance, a Likelihood-ratio test rejects the null hypothesis of using a Cobb-Douglas functional form.

conditions are tested at estimation time. The economic model is specified in Equation 1, where utility and time subscripts are omitted for clarity.

$$\ln TCost = \alpha + \beta_T T + \frac{1}{2}\beta_{T2} T^2 + \sum_m \alpha_m \ln Y_m + \sum_i \alpha_i \ln P_i + \frac{1}{2}\sum_m \sum_n \gamma_{mn} \ln Y_m \ln Y_n$$
$$+ \frac{1}{2}\sum_i \sum_j \gamma_{ij} \ln P_i \ln P_j + \sum_m \sum_i \gamma_{mi} \ln Y_m \ln P_i; \quad m, n = 1,2; i, j = 1..3; \tag{1}$$

In Equation 1, TCost is the total annual operating costs as defined previously; α is the intercept common to all firms; T is the time trend capturing possible technological changes occurring in the sector during the ten-year period, and T^2 represents the rate at which these changes may have happened; Y_m is the vector of the *m* outputs already identified: volume of water produced (Y_1), and volume of water loss (Y_3); P_i is the vector of **i** input prices previously identified as price of direct labor (P_1), price of indirect labor (P_2), and price of capital (P_3); the α's and γ's are parameters to be estimated. Table 1 depicts summary statistics for outputs and input factor prices.

According to Equation 1, costs are represented by a single functional form common to all the utilities with no systematic differences in technology among them. To account for technology differences due to the location of utilities in the different regions of the country, dummy variables are included in the model. The dummies are defined as R1, representing the *Mountain*, which takes a value equal to one if the utility is located in that region, zero otherwise. R3 represents the *Coast*, and equals one if the utility is located in that region, zero otherwise. These dummies are interacted with the first order and squared output and input price coefficients. A time trend and a time trend squared are also included in the model as shown in Equation 2.

$$\alpha_m = a_0 + a_1 R1 + a_3 R3; \quad \gamma_{mm} = b_0 + b_1 R1 + b_3 R3; \quad for \quad m = 1,2$$
$$\alpha_i = c_0 + c_1 R1 + c_3 R3; \quad \gamma_{ii} = d_0 + d_1 R1 + d_3 R3; \quad for \quad i = 1,2 \tag{2}$$
$$\beta_T = e_0 + e_1 R1 + e_3 R3 \quad \beta_{T2} = f_0 + f_1 R1 + f_3 R3$$

The a's , b's, c's,d's,e's and f's are parameters to be estimated. They measure regional cost shifts as deviations from the intercept, the *forest* region (R2), which is taken as the base.[19] This functional specification allows for the examination of *regional differences* in the cost elasticity of

---

[19] Because R2 is omitted, a_0, b_0, c_0, d_0, e_0 and f_0 are the coefficients for R2. The coefficients corresponding to R1 are all of the form: $g_1 + g_0$; For R3, the coefficients are all of the form $g_3 + g_0$, where g represents a,b,c,d,e,f.

outputs and acknowledges the fact that input prices may differ among the regions. Including the time trend and time trend squared interacted with region allows for accurately modeling when there have been changes in technology that differ across regions.

Finally, the cost variance may increase as the size of the company increases, leading to the presence of heteroscedasticity. To acknowledge this problem, a set of dummies are included in the model to represent the sizes of the utilities, which follows the size classification utilized by SUNASS. The dummy *Size***b** equals one for large utilities, zero otherwise, and *Size***s** equals one for small utilities, zero otherwise. The medium size dummy is the one omitted. These variables are interacted with the regional dummies in the same faction as previous interacted variables.[20]

3.2 Estimation Procedure

The economic model specified in equations 1 and 2 can be used to estimate either average cost values or minimum cost values. According to Aigner, Lovell and Schmidt (1977), it is only after the pioneering work on efficiency by Farrell (1957) that the possibility of estimating a production frontier as opposed to an average production function is considered when examining the performance of a firm. In the case of the water industry of Peru, the interest in the utilities performance supports the appropriateness of estimating a frontier. When estimating cost using a translog functional form, input prices and outputs are standardized before estimation such that the translog represents a second order approximation of the cost technology. In this model, the estimated frontier represents the minimum cost attained by each utility over the ten year period, and the first order coefficients of the translog cost function are interpreted as elasticities for the average firm at minimum cost.

Conceptually, the industry production function is a frontier determined by the production process of those firms attaining maximum output with a set of inputs. Other firms in the industry fall short of the frontier due to the presence of production inefficiencies. Suppose a firm uses a set of inputs X= $(x_1, x_2, .., x_n) \in \Re^n_+$, available at prices P= $(p_1, p_2, \ldots\ldots, p_n) \in \Re^n_{++}$ to produce output $Y = (y_1, y_2, .., y_j) \in \Re^j_+$. A production frontier is represented by a function $f : \Re^n_+ x\ \Re^j_+ \rightarrow \Re_+$ with a parameter vector φ such that $f(X; \varphi)$ denotes the maximum output obtainable from input

[20] At a 0.001 level of significance, a Likelihood-radio test rejects the null hypothesis of not including these variables in the model. In addition, at a 0.05 level of significance a White-test could not reject the homoscedastic hypothesis.

vector X. If the firm minimizes costs, the dual of a production frontier is a cost frontier represented by a function $C : \Re^n_+ x\ \Re^n_{++}\ x\ \Re^j_+ \rightarrow \Re_+$ with the same parameter φ such that C (P, Y*;* φ) denotes the minimum expenditure on inputs required to produce output Y with input price vector P. Output in this regulated environment is exogeneous, so the minimization assumption leads to an input-oriented inefficiency approach.

In this framework, the cost frontier represents the "*best*" that can be achieved, so observed cost cannot be less than minimum possible cost, in other words, $P^TX \geq C(Y,P; \varphi)$. In particular, the potential each firm has to reduce costs holding output constant is referred as cost efficiency and it is represented by the ratio defined in Equation 3.

$$CE = \frac{C(Y, P; \varphi)}{(P^T X)} \leq 1 \tag{3}$$

Farrell (1957) defined cost efficiency (overall efficiency) as a product of technical and price (allocative) efficiency. Essentially, cost inefficiencies arise from a combination of either using a wrong mix of inputs given the prevailing relative prices (allocative inefficiency), or from purchasing a wrong amount of input quantities to produce a given level of output (technical inefficiency). As water utilities in Peru are municipally owned, the presence of a larger share of technical inefficiency relative to allocative inefficiency is plausible. The rationale is that more than the cost minimizing amount of labor might be used as members of the board of directors and managers may engage in creating unneeded job positions in exchange for political favors. Indeed, when considering the number of workers per one thousand connections, which is an efficiency indicator in the water sector worldwide, Peru has a value of five. In general, this indicator is larger for state/government own enterprises (above 5) when compared to the same indicator for private companies (2 or 3), which is considered optimal. In addition, given the regulatory contract in place, and the poor development of the sector, capital over usage is not an issue, as it may be in the electrical sector.

Overall, predicting the shares of technical and allocative efficiency in this sector may be challenging without an appropriate analytical tool. Moreover, whether it is better to detect them separately depends on the objective of the cost analysis. This identification is possible in a cost frontier model only if input cost shares data are available and included into a simultaneous-equation model.[21] The interest of the current analysis is on overall efficiency with the benefit that the analysis comprises only a single cost equation, which simplifies the estimation process.

[21] For a discussion, see Kumbhakar and Lovell (2000), pages 131-183.

Following Atkinson and Cornwell (1994), the general form for a cost frontier is specified in Equation 4 (the time subscript and other variables already identified have been omitted for simplicity):

$$C_i(\mathrm{Y_i}, \mathrm{P_i}/u_i) = \min[(\mathrm{P_i}'/u_i)(u_i\mathrm{X_i}) \mid f(u_i\mathrm{X_i}) = \mathrm{Y_i}] = (1/u_i)C_i(\mathrm{Y_i}, \mathrm{P_i}) \quad (4)$$

In Equation 4, u_i is a parameter which measures the extent to which minimal cost differs from actual cost, the cost inefficiency ($0 < u_i \leq 1$) . The last equality in Equation 4 follows from the fact that a cost function is linearly homogeneous in input prices. In general terms, applying natural logs to Equation 4 yields the estimated cost model specified in Equation 5.

$$\ln C_i(Y_i, P_i / u_i) = \ln(1/u_i) + \ln C_i(Y_i, P_i) \Rightarrow -\ln u_i + \ln C_i(Y_i, P_i) \quad (5)$$

Once a frontier is selected, there is a need to specify whether this frontier is stochastic[22] or deterministic depending on the assumptions about the disturbance term.[23] Using a deterministic frontier implies that cost deviations from the frontier are considered to be pure inefficiency, as specified in Equation 3. Consequently, inefficiency may be confounded with noise or model misspecification which leads to biased estimates. A stochastic frontier framework imposes distributional assumptions on the disturbance term as having two components: a systematic and symmetric component (***v***), and inefficiency (***u).*** Inefficiency represents unobserved factors which are in control of the firm, such as those coming from the will and effort of the producer and his employees.[24] The idiosyncratic error term (***v)*** is assumed to be independently and identically distributed following a normal distribution with zero mean, and independent of the explanatory variables. A stochastic frontier is a sensible choice for this analysis because it acknowledges the presence of data typing and reporting errors. For the stochastic cost frontier, the definition of cost efficiency is specified in Equation 6.

$$CE = \frac{\mathrm{C(Y, P;}\varphi).(\exp\{v\})}{(\mathrm{P^T X})} \quad (6)$$

With a ten-year period of observed data, it is plausible to think that the inefficiency behavior of the utilities may have changed over time. To model this possibility, time is included

[22] Aigner, Lovell, and Schmidt (1977) were the first to introduce a stochastic frontier approach, in parallel with Meeusen and Van den Broeck (1977). The stochastic approach was extended to the panel data case by Pitt and Lee (1981) and Schmidt and Sickles (1984).

[23] For a discussion, see Murillo-Zamorano (2004) and Kumbhakar and Lovell(2000)

[24] Jensen (2005) compares both approaches for cross-section models and finds differences on estimated inefficiencies depending upon the size of the data set, technology functional form, and the objective of estimation: firms' efficiency levels or ranking.

in the estimation process following Battese and Coelli's (1992) functional form specification for the inefficiency component as stated in Equation 7.

$$u_{it} = u_i \exp[-\eta(t-T)];\;\; u_i \; iid \sim N(\mu, \sigma^2) \qquad (7)$$

In Equation 7, the u_i are assumed to be independent and identically distributed non-negative truncations of a Normal distribution with parameters μ and σ^2; eta (η) is a scalar parameter to be estimated; t represents each time period within the T total number of periods. With this specification, as t increases, u_{it} decreases if η is higher than zero, remains constant if η equals zero, and increases if η is less than zero. This specification assumes that all firms follow the same trend. Although restrictive, still this is a plausible assumption for the water sector in Peru because, once region and size have been controlled for, all utilities deliver service under the same economic, social and political circumstances, following the same regulatory rules.

4. EMPIRICAL RESULTS

4.1 Regularity of the Cost Function

The model specified in Equations 1 and 2 is estimated as a stochastic panel cost frontier using Maximum Likelihood estimation. Estimation proceeds using a balanced data set of 43 utilities and 10 years, comprising a total of 430 observations. Table 2 shows the results of the estimation. A well-behaved cost function is concave in input prices and non-decreasing in outputs. Assuming the cost function is twice continuously differentiable, a necessary and sufficient condition for it to be concave in prices is that the matrix of second order partial derivatives of the cost function with respect to prices is negative semi-definite. In the case of the translog functional form, for this to hold it is necessary to impose symmetry on the parameters of interacted price terms. This is accomplished by applying $\gamma_{ij} = \gamma_{ji}$ for all $i \neq j$. In addition, following Diewert and Wales (1987), the price shares need to be positive over the price domain. Price shares are found to be positive across all regions at mean data values.

A cost function must be homogeneous of degree one in prices to correspond to a well-behaved production function. This implies that for a fixed level of output, total cost must increase proportionally when all prices increase proportionally. This is accomplished by imposing the restrictions specified in Equation 8.

$$\sum_i \alpha_i = 1, \quad \sum_i \gamma_{Yi} = 0, \quad \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = \sum_i \sum_j \gamma_{ij} = 0 \qquad (8)$$

Imposing these restrictions is equivalent to normalizing prices and total cost by one of the input prices. Total costs, price of direct labor and price of capital are divided by the price of indirect workers (P_2).[25]

A cost function corresponds to a homothetic production technology if and only if the cost function can be written as a separable function in output and factor prices. A homogeneous technology is a special case of a homothetic technology when the elasticity of cost with respect to output is constant. Following Christensen and Greene (1976) and Diewert (1974) the homotheticity and homogeneity conditions are tested using a Likelihood Ratio test after imposing the restrictions specified in Equations 9 and 10 respectively.

Homotheticity requires: $\gamma_{Yi} = 0$ (9)

Homogeneity in outputs requires: $\gamma_{Yi} = 0; \quad \gamma_{YY} = 0$ (10)

A Likelihood Ratio test at a 0.001 level of statistical significance could not reject the null hypothesis of homotheticity. The production function being homothetic implies that the input mix is constant with scale. The homogeneity hypothesis is rejected after a Likelihood Ratio test at a 0.001 level of statistical significance. This implies that returns to scale vary with the scale of the firm and the production mix. In the following sections the homothetic specification is used when reporting results.

4.2 Efficiency, Technology Changes, and Size

For the frontier estimator the likelihood function is expressed in terms of the variance parameters, $\sigma^2 = \sigma_V^2 + \sigma_u^2$ and gamma, $\gamma = \sigma_u^2/\sigma^2$. The closer gamma is to one, the more it is inefficiency, not noise, that accounts for the variance in the disturbance term. A large value for gamma (0.96) indicates that the variance of the error term is mainly explained by unobserved inefficiency. This parallels the fact that the inefficiency mean value is statistically significant at a 99% level. The coefficient for η is very small and not statistically significant, indicating that firms in this industry have not changed their efficiency behavior, or at least not following the time change functional form imposed to the model.[26]

---

[25] The selection of the price for normalization does not alter the results.

[26] The null hypothesis of using Ordinary Least Squares (OLS) estimation is rejected at 0.01 level of statistical significance using a Likelihood-ratio test.

Inefficiency statistics are shown in table 3 and 4 using cost reduction as a measure of cost inefficiency. Cost reduction is the extent to which utilities would need to reduce costs if they were to behave at 100% efficiency. Cost reduction is calculated as minus the natural log of the estimated cost efficiency via E (u_{it} | e_{it})) divided by actual costs. When looking at cost reduction values on a year by year basis by region as depicted in Table 3, it is evident that utilities on the coast (R3) are more efficient than utilities in other regions. The overall cost reduction for these utilities is only 11.5% when compared to 14% for utilities in the forest (R2) and 16% for those in the mountains (R1). Table 4 shows cost inefficiency statistics by size of the company and location on a year by year basis. Overall, large utilities are the most efficient. Cost reductions for large utilities are approximately 9%, and for medium and small companies they are 13% and 16%.

Overall, the coefficients for the time variables are very small which agrees with the low level of technological changes in the sector. For firms in the forest (R2) and mountains (R1), the time coefficient is statistically different from zero indicating that the cost frontier has shifted over time in these regions. In the case of the forest (R2), the technology has shifted towards the origin indicating a cost decrease, at an increasing rate, and the opposite is true for the mountains (R1), indicating an increase in costs at a decreasing rate. When the size of the utilities are considered, results indicate 1.48% lower costs for small firms in the forest (R2) when compared to the costs of small utilities in other regions.

4.3 Output Variables and Economies of Scale

First order output coefficients are interpreted as output cost elasticities in the translog functional form. The regional coefficients for water produced (y1) are large and positive as expected. However, only the one for the forest (R2) is statistically different from zero with 90% confidence. This indicates that, as volume of water produced increases by 1%, the impact on costs is 2% larger for utilities in the forest than for utilities in other regions. However, given the negative and statistically significant coefficient on the water produced squared term, this effect is decreasing at a 0.2% rate.

With respect to water lost the coefficient for the interacted term (y1y3) is positive and statistically significant at a 95% confidence level. This finding supports the hypothesis that water lost is being jointly produced in this industry. The positive coefficient indicates diseconomies of scope for the joint production. A 1% increase of water production while allowing water to be lost implies an increase in the cost of the utility of 0.10%. Under the framework of this analysis, it is plausible to consider the 0.10% the cost of allowing non-authorized connections in the system. Or it might reflect the cost of capturing municipal votes from the poorest segment of the population in the area of service. Alternatively, the 0.10% may also represent the opportunity cost of repairing network leaks and exerting stricter network maintenance and control.

From an economic point of view, this industry is a natural monopoly, so finding economies of scale is not surprising. Following Christensen and Greene (1976), economies of scale are defined in terms of the relationship between total cost and output along the expansion path where input prices are constant and costs are minimized at every level of output. The elasticity of cost with respect to output (ε) [27] is generally used in the calculation of economies of scale. Following Baumol (1976), and Panzar and Willig (1977), a local measure of overall scale economies for a multi-product firm is defined in Equation 11.

$$ES = 1 - \sum_i \varepsilon_i \quad ; \quad \text{where} \quad \varepsilon_i = \frac{\partial \ln TC}{\partial \ln Y_i} \quad \text{for i=1..3} \tag{11}$$

In Equation 11, a positive (negative) ES implies economies (diseconomies) of scale. The calculation is performed using mean data values, and firms are classified by size and region when reporting the results in Table 5. Findings show the presence of economies of scale across all regions, indicating that average costs are declining in the industry even for large values of output, as it was expected.

Overall, economies of scale are smallest on the coast (0.58) and largest in the forest (0.81), independently of the size of the utility. These findings suggest that firms in the forest (R2) should be considered first if the regulatory agency considers consolidation of utilities while opening the sector to private participation.

4.4 Input Factors and Price Elasticity

The effect on costs of a 1% rise in direct labor prices (p1) is 0.01% smaller on the coast (R3) than in the forest (R2). It is 0.001% smaller in the mountains (R1) than in the forest. In parallel, the effect on costs of a 1% rise in the price of capital (p3) is 0.02% larger on the coast (R3) than in the forest or mountains. Overall, these results indicate over-usage of debt and under-usage of direct labor on the coast (R3) relative to their usage in other regions.

The coefficient for the interacted prices of labor and capital (p1p3) is positive and statistically significant at a 0.01 level, indicating that these inputs are substitutes. Substitutability

---

[27] The elasticity of cost is defined as the proportional increase in costs resulting from a proportional increase in the level of output.

of labor and capital is not a surprising result in this sector. As the price of capital (investment in network) increases, network expansion is followed by an increase in the number of workers. Conversely, an increase in the price of labor is followed by a network increase. As symmetry was imposed in the translog functional form, further investigation on this substitutability is performed utilizing Morishima elasticities.[28] Table 6 shows calculations for both labor-capital and capital-labor elasticities evaluated at mean data values.

In interpreting these elasticities, the larger the value of the positive elasticity, the easier is the substitution of the input factors. Thus, it is easier to substitute labor for capital in the forest (R2) than in the other regions. In other words, a change in one of the input prices produces a quicker reaction in the input mix for utilities in the forest than in those at other regions. For utilities on the coast (R3), the inputs labor and capital are complements given the negative elasticity, but they are substitutes when considering capital and labor. However, the magnitudes of both elasticities are very small. One way to explain this result is that very small elasticities indicate that factor proportions are held on to so tightly that they must be needed in relatively fixed proportions.

4.5 Testing the Results

The econometric model is based on a crucial assumption: input factor prices, outputs, region, and size are not correlated to the residual. If they are, results of estimation are biased. The discussion on how to explain inefficiency, found in the empirical literature, relates to the difficulties of performing a statistical test for this assumption. [29] The critical aspect in this discussion, and in performing the test, is that inefficiency is unobserved, so its estimation is challenging. The Battesse and Coelli (1995) model is selected to perform the test as it is classified by Kumbhakar and Lovell (2000) as a "recent approach" to explain inefficiency overcoming some of the difficulties of previous approaches. [30] This model implies a "direct" explanation of inefficiency as depicted in Equation 12.

$$\ln C_i = \ln f(Y, P; \beta) + v_i + u_i \quad ; \quad u_i = \gamma' Z + \varepsilon_i \geq 0 \qquad (12)$$

[28] The Morishima partial elasticity of substitution is a measure of elasticity of substitution utilized in the multi input case and proposed by Michio Morishima (1967).

[29] For details see Kumbhakar and Lovell (2000), Coelli, Perelman and Romano (1999), Greene (2005a), (2005b) and (2004).

[30] The use of this model is suggested by an anonymous referee.

In this model, firms' inefficiency, u_i is explained by the vector of variables Z, and γ is a vector of parameters associated to these variables. In this specification, the variables comprising Z, in this particular case region and size, enter the inefficiency equation but not the cost technology specification. Thus, region and size are supposed to explain inefficiency as if firms have the choice of selecting the region where to provide services and have control on its size. Estimation results show large t values for all the coefficients but the coefficients for region and size are not statistical significant. The lack of statistical significance indicates that region and size do not explain inefficiency, which agrees with the exogenous assumptions of the original model and indicates that its estimation results are reliable.

5. CONCLUDING OBSERVATIONS

In 2006, the performance of the water sector was a main concern for the water regulatory agency of Peru. Findings from this study reveal important points regarding the performance of utilities in this industry, which may help the agency in the decision-making process while opening the sector to private capital participation. For the period 1996 to 2005, economies of scale are found across all regions, as the sector is a natural monopoly. Consequently, the aggregation of utilities in this industry is advised as the participating companies will benefit from larger cost savings. However, utilities in the forest (R2) show larger economies of scale than do utilities in other regions. Thus, service providers in the forest should be the first to be considered for consolidation if such an action is in the set of policy decisions considered by the regulatory agency.

The study finds utilities on the coast (R3) to be more efficient than utilities in other regions, except for utilities of small size: these are more efficient in the forest. When considering size, larger utilities are more cost efficient than utilities of other sizes independently of the region.

Findings also indicate that cost technology has changed over time for utilities in the forest and in the mountains, but this change is cost-decreasing for utilities in the forest whereas it is cost increasing for utilities in the mountains. Particularly, findings indicate increasing volume of water produced is more costly for utilities in the forest, a critical aspect to be considered when expanding service.

Regarding the price of input factors, regional differences are important for utilities on the coast only. The effect of direct labor prices on utility costs is smaller on the coast than in the other regions. On the other hand, the effects of capital prices are higher on the coast than in the forest or mountains.

Utilities on the coast need capital and labor in fixed proportions, reflecting higher rigidities when substituting input factors compared to the easier substitutability of input factors found in utilities located in the forest. This result may reflect the fact that utilities on the coast are of a larger size than those in the forest.

Finally, results related to the set of outputs reveal that costs increase by 0.1% with a 1% increase of water produced while allowing water to be lost. Under the framework of this analysis, this cost increase represents the price paid by firms for allowing non-authorized connections in the network and, presumably, for securing the votes of the population being served with water that is not billed. The effect of political interference in the provision of water in this country seems to be an interesting topic for future research.

Table 1. Summary statistics for outputs and input prices from 1996 to 2005 by size

Variable		Mean	StDv	Min	Max
Volprod – Millions m3 (Y1)	Large	116	209	11.9	705
	Medium	9.9	4.3	2.7	22
	Small	2.7	1.5	0.3	6.3
Volwaterloss- Millions m3 (Y3)	Large	49	8.4	0.7	317
	Medium	5	2.8	0.2	10
	Small	1.4	1	0.03	4.7
PriceDwork - $/worker (P1)	Large	10.9	4.3	3.3	23
	Medium	7.1	2.4	1.8	19
	Small	5.9	2.6	0.4	18.3
PriceIwork - $/worker (P2)	Large	6.2	5.5	0.39	21
	Medium	4.8	4.2	.63	23
	Small	4	3.3	0.5	32
PriceofCapital - $/km (P3)	Large	4.7	2.7	1.08	11
	Medium	3.2	2.4	0.32	14
	Small	2.9	2.7	0.1	13

Table 2. Estimation results for the translog cost function

Forest		Mountains		Coast	
Var	Coeff	Var	Coeff	Var	Coeff
t	-0.04* (0.021)	tR1	0.065** (0.026)	tR3	0.011 (0.026)
t2	0.005*** (0.001)	t2R1	-0.006*** (0.002)	t2R3	-0.002 (0.002)
*Size***b**	0.374 (0.727)			*Size***b** R3	0.971 (0.787)
*Size***s**	-1.48*** (0.515)	*Size***s** R1	0.049 (0.55)	*Size***s** R3	0.812 (0.686)
y1	2.08* (1.21)	y1R1	-0.997 (0.969)	y1R3	-1.046 (0.948)
y1y1	-0.212** (0.104)	y1y1R1	0.08 (0.071)	y1y1R3	0.1 (0.069)
y3	-0.558 (0.897)	y3R1	0.872 (0.986)	y3R3	0.698 (0.97)
y3y3	-0.062 (0.075)	y3y3R1	-0.07 (0.073)	y3y3R3	-0.061 (0.072)
y1y3	0.095** (0.046)				
p1	0.001 (0.001)	p1R1	-0.002 (0.073)	p1R3	-0.012*** (0.004)
p3	-0.001 (0.002)	p3R1	0.001 (0.002)	p3R3	0.024*** (0.008)
p1p1	-0.012 (0.024)	p1p1R1	-0.045 (0.052)	p1p1R3	-0.027 (0.03)
p3p3	0.117*** (0.034)	p3p3R1	0.031 (0.041)	p3p3R3	0.025 (0.039)
p1p3	0.03*** (0.009)				
				Gamma	0.96
intercept	-6.45 (4.84)	**eta**	-0.0003 (0.005)	**mu (*u*)**	0.95*** (0.32)

Dependent variable: Ln (Total Operating Costs); Confidence levels: *** 99%; ** 95%; * 90%
Data set: 430 observations = 43 groups, 10 years per group; Standard Errors in parenthesis. Loglikelihood=199.832

Table 3. Cost reduction statistics from 1996 to 2005 by region

In %	Mountains (R1)			Forest(R2)			Coast(R3)		
year	**Mean**	Max	Min	**Mean**	Max	Min	**Mean**	Max	Min
1996	16.38	26.07	1.05	13.66	25.64	5.01	11.40	20.65	1.22
1997	16.33	26.43	.88	13.81	25.67	4.96	11.44	20.55	1.24
1998	16.10	27.38	.87	13.43	24.99	4.99	11.43	20.59	1.23
1999	16.20	27.58	.88	13.60	25.71	4.81	11.56	20.54	1.24
2000	16.14	27.58	.90	13.64	25.90	4.76	11.60	20.50	1.23
2001	16.01	28.32	.89	13.70	26.12	4.81	1159	20.37	1.25
2002	16.05	28.24	.88	13.56	25.49	4.81	1154	20.24	1.24
2003	16.00	27.78	.89	13.58	25.38	4.82	11.48	20.22	1.24
2004	15.81	27.91	.88	13.39	25.03	4.75	11.41	20.14	1.24
2005	15.77	27.86	.90	13.29	25.24	4.72	11.51	19.97	1.23
Total	**16.08**	**28.32**	**.87**	**13.57**	**26.12**	**4.72**	**11.5**	**20.65**	**1.22**

Table 4. Cost inefficiency mean values from 1996 to 2005 by utility size and region of location

Cost Reduction in %	(Mean)	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Large	Mountains (R1)	-	-	-	-	-	-	-	-	-	-
	Forest (R2)	11.18	10.86	10.73	11.19	11.05	11.37	11.06	11.23	11.22	11.08
	Coast (R3)	09.09	09.05	09.07	09.09	09.10	09.11	09.10	09.10	09.08	09.03
	All	09.35	09.28	09.27	.09.35	09.35	09.39	09.34	09.36	09.34	09.29
Medium	Mountains (R1)	14.52	14.39	14.20	14.29	14.23	14.26	14.12	14.06	13.97	13.92
	Forest (R2)	12.07	12.20	12.05	12.25	11.97	12.10	11.89	12.02	11.92	11.79
	Coast (R3)	12.25	12.25	12.28	12.52	12.56	12.56	12.46	12.43	12.32	12.58
	All	13.31	1326	13.16	13.33	13.29	13.31	13.19	13.16	13.06	13.13
Small	Mountains (R1)	18.75	18.81	18.54	18.64	18.58	18.25	18.53	18.49	18.17	18.14
	Forest (R2)	14.60	14.83	14.33	14.44	14.61	14.61	14.52	14.48	14.24	14.14
	Coast (R3)	14.50	14.83	14.64	14.72	14.80	14.76	14.73	14.50	14.42	14.39
	All	16.40	16.57	16.23	16.33	16.38	16.23	16.31	16.24	15.99	15.94

Table 5. Economies of scale

	Large	Medium	Small	All
Mountains (R1)	-	0.79	0.67	0.74
Forest (R2)	0.94	0.89	0.75	0.81
Coast (R3)	0.64	0.22	0.51	0.58

Table 6. Morishima elasticities

	Labor – Capital	Capital - Labor
Mountains (R1)	0.4	1.0
Forest (R2)	2.4	1.7
Coast (R3)	-0.7	0.6

6. REFERENCES

Aigner D, Lovell C, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. J Econom 6: 21–37

Antonioli B, Filippini M (2001) The use of variable cost function in the regulation of the Italian water industry. Utilities Policy 10:181–187

Atkinson S, Cornwell C (1994) Estimation of output and input technical efficiency using a flexible functional form and panel data. Inter Econ Rev 35: 245–255

Aubert C, Reynaud A (2005) The impact of regulation on cost efficiency, an empirical analysis of Wisconsin Water Utilities. J Prod Anal 23:383–409

Baumol W (1976) Scale economies, average cost, and the profitability of marginal cost pricing. Essays in Urban Economics and Public Finance, R.E. Grieson, Lexington, Mass. 43–57

Battese G, Coelli T (1992) Frontier production functions, technical efficiency and panel data: with application to Paddy Farmers in India. J Prod Anal 3:153–169

Battese G, Coelli T (1995) A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics 20:325–332

Coelli, T, Perelman, S., & Romano, E. (1999) Accounting for environmental influences in stochastic frontier models: with application to international airlines. J Prod Anal 11:251-273

Corton ML (2003) Benchmarking in the Latin American water sector: the case of Peru. Utilities Policy 11:133–142

Christensen L, Greene W (1976) Economies of scale in U.S. electric power generation. J Pol Econ 84(4): 655–676

Diewert WE (1974) Applications of duality theory. In Frontiers of Quantitative Economics Amsterdam: North-Holland

Diewert WE, Wales TJ (1987) Flexible functional forms and global curvature conditions. Econometrica 55(1):43–68

Farsi M., Filippini, M., & Greene, W. (2005) Efficiency measurement in network industries: application to the Swiss railway companies. J Reg Econom 28(1):69-90.

Farrell M (1957) The measurement of productive efficiency. J Royal Stat Soc, series-A 120(3):253–282

Garcia S, Thomas A (2001) The structure of municipal water supply costs: application to a panel of French local communities. J Prod Anal 16: 5–29

Greene, W. (2005a) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. J Econometrics 126:269-303.

Greene, W. (2005b) Fixed and random effects in stochastic frontier models. J Prod Anal 23:7-32.

Greene, W. (2004) Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems. Econometrics and Health Economics 13:959-980.

Hattori T (2002) Relative performance of US and Japanese electricity distribution: an application of stochastic frontier analysis. J Prod Anal 18:269–284

Hattori T, Jamasb T, Pollitt M (2005) Electricity distribution in the UK and Japan: a comparative efficiency analysis 1985–1998. The Energy Journal 26(2):23–47

Jamasb T, Pollitt M (2003) International benchmarking and regulation, an application to European electricity distribution utilities. Energy Policy 31(15):1609–1623

Jensen U (2005) Misspecification preferred: the sensitivity of inefficiency rankings. J Prod Anal 23:223–244

Kingdom B, Liemberger R, Marin P (2006) The challenge of reducing non-revenue water (NRW) in developing countries, how the private sector can help: a look at performance based serving contracting. World Bank Water Supply and Sanitation Sector Board Discussion Paper Series, 8, December 2006

Kumbhakar SC, Lovell CAK (2000) Stochastic frontier analysis. Cambridge University Press, Cambridge

Lowry M, Getachew L, Hovde D (2005) Econometric benchmarking of cost performance: the case of U.S. power distributors. The Energy Journal 26(3):75–92

Murillo-Zamorano, L. (2004). Economic efficiency and frontier techniques. J Economic Surveys 18(1): 33–45.

Meeusen W, Van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. Inter Econ Rev 18:435–444

Morishima, M. (1967) A Few Suggestions on the Theory of Elasticity. Keizai Hyoron

Neuberg L (1977) Two issues in the municipal ownership of electric power distribution. Bell J Econ 8(1):303–323

Panzar JC, Willig R (1977) Economies of scale in multi-output production. Quart J Econ 91(3):481–493

Pitt M, Lee L (1981) The measurement sources of technical inefficiency in the Indonesian weaving industry. J Dev Econ 9:43–64

Renzetti S (1999) Municipal Water Supply and Sewage Treatment Costs, Prices and Distortions Canadian. J Econ 32(3):688–704

Rossi Martin (2001) Technical change and efficiency measures: the post privatization in the gas distribution sector in Argentina. Energy Economics 23:295–304

Saal D, Parker D (2000) The impact of privatization and regulation on the water and sewerage industry in England and Wales: a Translog cost function model. Managerial and Decision Economics 21(6):253–268

Saal D, Parker D (2004) Productivity and price performance in the privatized water and sewerage companies of England and Wales. J Reg Econ 20(1):61–90

Schmidt P, Stickles R (1984) Production frontiers and panel data. J Bus Econ Stat 4: 367–374

SUNASS (2006) Infraestructura de agua potable y alcantarillado urbano en el Peru: un reto pendiente. Copyright SUNASS, May 2008. http://www.sunass.gob.pe/publicaciones.php (last visit: 02/24/2010).

Tamayo G, Barrentes R, Conterno E, Bustamante E (1999) Reform efforts and low–level equilibrium in the Peruvian water sector. In Savedoff W, Spiller P, Spilled Water: Institutional Commitment in the Provision of water Services. Inter-American Development Bank, Washington D.C. 89–133