Consistency in Performance Rankings: The Peru Water Sector

By Sanford Berg and Chen Lin[1]

April 11, 2006 (revised June 26, 2007). Forthcoming in *Applied Economics*

Regulators can utilize a number of alternative methodologies for comparing firm efficiency, but these approaches need to be robust to be accepted by stakeholders. This study evaluates the consistency of water utility performance rankings for Peruvian water utilities. The results indicate that Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis yield similar rankings in this case. In addition, the techniques have comparable success for identifying the best and worst performing utilities. However, these rankings based on sophisticated statistical techniques are not highly correlated with those developed by the Peruvian water regulator (SUNASS). This result does not invalidate the performance rankings obtained by the regulator, since those rankings are based on more dimensions of utility performance. However, they illustrate the importance of developing sound techniques for identifying weak utilities. Improvements in sector performance require that benchmarking be given greater attention than in the past.

Keywords: Water Utility Performance, Data Envelopment Analysis, Stochastic Frontier Analysis

[1] Berg (sberg@ufl.edu) is Distinguished Service Professor of Economics and PURC Director of Water Studies,; Lin (chen.lin@cba.ufl.edu) is Assistant Professor, Lingnan University, Hong Kong and PURC Research Associate, University of Florida.

Consistency in Performance Rankings: The Peru Water Sector

By Sanford Berg and Chen Lin
University of Florida and Lingnan University, Hong Kong

1. Introduction

The purpose of this study is to demonstrate the benefit of using alternative methodologies for making performance comparisons across water utilities. Others have demonstrated the strengths and limitations of various methodologies in addressing issues ranging from the effects of ownership on efficiency, impacts of incentives, and measurements of productivity change.[2] The present study utilizes Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) to rank relative performance within Peru, where all of the water utilities are publicly owned.[3]

Since the Peruvian water regulator (Superintendencia Nacional de Servicios de Saneamiento—SUNASS) publishes a ranking of water utilities (along with supporting data), this nation serves as a good case study of utility performance evaluation. Given the difficulty of creating strong incentives for cost containment and service expansion for publicly owned water utilities, yardstick comparisons represent a key tool for regulators in developing countries. Widely available scorecards of relative utility performance can put public pressure on under-performing firms. This study utilizes both parametric and non-parametric benchmarking methods to obtain comparisons across water utilities.

The collection and analysis of water utility data represents an important tool for documenting past performance, establishing baselines for gauging improvements, and making comparisons across service providers. In the water sector particularly, valid comparisons can contribute to improved performance. Rankings can inform policymakers, the providers of investment funds, and customers regarding the cost

[2] Recent studies have focused on the relative performance of public and private water utilities in the United States (Bhattacharyya, Parker, and Raffiee 1994; Wallsten and Kosec 2005), the United Kingdom (Hunt and Lynk 1995; Cubbin, and Tzanidakis 1998; Ashton 2000; Saal and Parker 2001), Asia (Estache and Rossi 2002), Latin America (Clarke, Kosec, and Wallsten 2004; Tupper and Resende 2004), and Africa (Estache and Kouassi 2002; Kirkpatrick, Parker, and Zhang 2004). For a more detailed and comprehensive literature review, readers are referred to the recent work by Estache et al. (2005), who survey recent productivity and efficiency literatures in infrastructure industries (energy, ports, railways, roads, telecommunications and water/sewerage) in developing countries.

[3] Researchers have used a combination of DEA and SFA (or other methodologies) to assess technical efficiency in several sectors. For example, Mizala, Romaguera, and Farren, D. (2002) analyze the technical efficiency of schools in Chile, finding that the two methods give similar rankings for schools. Latruffe, Balcombe, Davidova, and Zawalinska.(2004) report the determinants of technical efficiency for crop and livestock farms in Poland (DEA-second stage analysis is used to check the robustness of the SFA results). Also, Fiordelisi and Molyneux (2004) examined efficiency and productivity in the Italian factoring industry using DEA and the Malmquist Index.

effectiveness of different service providers. Utility managers are concerned about how such rankings reflect on their capabilities, so they may wish to limit information available to regulators, avoiding public scrutiny in the press. The audiences for yardstick comparisons have different degrees of expertise and interest when it comes to evaluating water utilities; nevertheless, those concerned with the production efficiency will want rankings that reflect reality. If the criterion of consistency is not met, these groups cannot be confident that the relative performance indicators are meaningful. However, when alternative methodologies yield broadly similar rankings, stakeholders are less likely to engage in acrimonious high-stakes disputes. Despite the topic's importance, only a few papers examine the sensitivity of efficiency and rankings based on different methodologies. Cubbin and Tzanidakis (1998) compare the rankings based on OLS and DEA using a sample of 29 U.K. water utilities in year 1994. They found that the rankings based on the OLS model are different from the ones based on DEA models. In this study, we conduct a systematic check of the sensitivity of efficiency and rankings based on different methodologies, including OLS, DEA (CRS, VRS, and Super-efficiency), SFA distance function model and the evaluation system employed by the regulator.

The present study is directed to economists and policy analysts responsible for producing, evaluating, and interpreting benchmarking studies. For example, Tupper and Resende (2004) use DEA supplemented by econometric analyses to provide efficiency scores for twenty Brazilian state water and sewage companies. Using a different methodology, Corton (2003) studied forty-four Peruvian water utilities, estimated a cost function, and ranked firms in terms of their performance relative to that predicted by the estimated model. In a paper that focused on service quality in Peru, Lin (2005) utilizes SFA to demonstrate the importance of incorporating chlorine tests, accounted for water, service continuity (hours per day), and coverage in benchmarking models. The present paper only includes two aspects of service quality (continuity and coverage) in order to provide more comprehensive comparisons across a broader range of alternative model specifications (both SFA and DEA).

To be of use to regulators who implement infrastructure policy, and to be accepted by other stakeholders, performance comparisons must be robust to promote confidence that the performance rankings do indeed reflect managerial skill rather than accidents of geography or history. Typically, the regulatory commission reviews studies and establishes performance incentives to achieve policy objectives. Without confidence in the scores and relative rankings, those responsible for creating incentives will not risk their credibility by instituting rewards or applying penalties. Regulators will be unwilling to apply incentives based on performance unless they are very confident that the rankings can survive challenges.

Thus, relative and absolute rankings can become catalysts for improved cost containment and service quality if stakeholders have confidence in the analysis. *Relative* rankings allow stakeholders to compare the performance of utilities in comparable situations. Here, the key problem is how to select firms that are truly similar to one another. Analysts want the relative ranking to reflect managerial decisions rather than the unique characteristics of service territories beyond managers' control, including topography, hydrology, and customer density. In addition, history matters: current managers have inherited utility systems that reflect a set of political and economic (including regulatory) decisions made by others. Thus, performance improvements over time also need to be taken into consideration.

Absolute comparisons are also necessary, since the weakest performer in one group might have much better performance than the best firms in another group of comparable firms (say, those in another country at a similar stage of development). Comparisons are valid so long as the results do indeed tell us whether particular firms are performing below their potential. Therefore, the problem of consistency requires a reasonable, reliable, and stable scorecard that takes into account the political and economic environment of different countries, the specific targets set by the regulators, the availability of the data, and other factors.

Two recent studies of Peru note the problems faced within the sector. Torero and Pasco-Font (2003, p. 283) point out that the water sector has made some progress, but ". . . water is available for limited periods of time, and its quality is poor. Consequently households incur additional costs in efforts to improve quality and to ensure access." Corton (2003) agrees that the water sector in Peru is characterized by serious problems, including inadequate system maintenance, a high level of unaccounted-for water, excess staff, low metering rates, and low water quality. The lack of readily available comparative data about quality, operation costs, quantity, and service coverage makes it hard for customers to exert pressure on the water companies to improve their performance. Performance measurement has been recognized as a way of introducing competitive pressures while motivating appropriate behavior by management.[4]

This study employs the parametric benchmarking model, SFA, and DEA, the non-parametric benchmarking model, to evaluate the cross-sectional efficiency of Peru's water industry. The rankings provided by SFA and DEA analysis are then compared to the SUNASS ranking and the regression model presented by Corton (2003). The latter approaches are presented first to illustrate the nature of the data used for benchmarking.

2. Past Studies: the SUNASS Ranking Method and a Regression Model

The Peruvian government created SUNASS in 1992 to regulate water and sanitation services. In late 1999, SUNASS established a benchmark system as a first step toward informing citizens and political leaders about the relative performance of the municipal utilities. The expectation was that low-efficiency companies would gradually reach the level of the most efficient companies in response to greater pressure to perform efficiently. SUNASS developed a system of efficiency indicators and a benchmarking scheme through internal working groups and guidance from a World Bank consulting team. The process consisted of identifying areas of efficiency affecting the interests of each group of stakeholders and then identifying

[4] As Smith (1990) observes, comparative data provide a mechanism whereby consumers can appraise the quality of local services. The benchmarking results can also help regulators decide the efficient cost and determine the appropriate price caps (Carrington et al. 2002). In the case of Peru, all the utilities are publicly-owned, so using higher X-factors (as a stick) for poorly performing firms may be inappropriate: citizens benefit from lower prices, but they are likely to experience lower service quality as the poorly managed utilities have reduced cash flows. Informed public opinion can exert pressure on managers via local elections. Perhaps devising managerial compensation schemes based on improvements in rankings offers greater promise (as a carrot).

indicators for these areas. Nine indicators were selected and grouped into four areas of efficiency.

1. *Quality of Service* includes three variables: compliance with the residual chlorine rule, continuity of service, and percentage of water receiving chemical treatment.
2. *Coverage of Service Attained* consists of two variables: water coverage and sewerage coverage.
3. *Management Efficiency* reflects three variables: operating efficiency (a combination of continuity of service and the volume of water produced to serve each person in a connection), percentage of connections with meters installed, and the ratio of bills not yet collected to the total number of bills (nonpayment).
4. *Managerial Finance Efficiency* is defined by the ratio of direct costs and other expenses to revenues.

The first two broad areas of efficiency are intended to represent the interests of society. The third reflects the companies' performance, and the fourth represents the citizen-owner's perspective. Each of the nine sub-indicators was assigned a weight of 1 as a first step in the benchmarking process. The emphasis on social concerns is evident in the greater number of indicators related to efficiency affecting society. Each indicator expressed as a percentage is multiplied by its weight and added to the percentage total per company. This total per company is divided by nine, the number of indicators. Finally, each company providing water and sanitation services (Empresas Proveedoras de Servicios or EPS) is ranked on the basis of points attained within a group, with groupings determined by number of service connections, as follows: 17 EPS are *small* with fewer than 10,000 connections; 20 EPS are *medium*-sized, with 10,000–40,000 connections;7 EPS are *big*, with 40,000–160,000 connections. Although SUNASS does evaluate SEDAPAL, the water utility of Lima, that EPS is not included in the present study since its high number of connections (nearly one million) would distort the results.

Each company reports its efficiency indicators within a master plan according to its financial and operational resources. This document is submitted annually to SUNASS for review and approval. Once the master plan is approved, it is viewed as involving mandatory targets. At the end of the year, the ranking based on actual data is calculated and published by SUNASS. Thus, the ranking represents a potentially contentious listing from the standpoint of municipal politics. However, SUNASS cannot (at present) implement formal rewards or penalties based on these rankings, so the comparisons are not widely distributed.

The dependent variable in Corton's regression model was operation cost; the independent variables are volume of water produced (Volprod), the length of mains (Length) measured in kilometers, and the number of districts administered by each company (Loc). Two dummies for region are included in the model: *RC*=1 if the company is located on the coast, *RR*=1 if the company is located in the mountains, 0 otherwise(forest)..The equation of the regression model is shown as:

$$\ln OperCost = \alpha_1 + \alpha_2 RC + \alpha_3 RR + \beta_1 \ln Volprod + \beta_2 \ln Length + \beta_3 \ln Loc + \mathrm{e} \qquad (1)$$

Level of efficiency= (Actual Operating Cost- Predicted Operating Cost) / (Predicted Operating Cost)

The ranking results are quite different in the SUNASS evaluation and the Corton regression model. Furthermore, both models have limitations. In the SUNASS model, an equal weight of 1 is assigned to each of the nine indicators. This weighting does not differentiate among the different performance dimensions: equal weighting needs some justification. More importantly, most indicators in the SUNASS scheme lack input-output causative relationships. Only the indicator for managerial efficiency considers the cost issue. The basic definition of productivity is output/inputs. Therefore, the benchmarking scheme in the Peruvian water sector is actually an output-oriented benchmarking scheme, not a productivity benchmarking scheme. The Corton model focuses on the cost efficiency of the company. It does not consider other important factors such as input price, quality of service and coverage of service. In addition, an inherent problem of regression analysis is that it requires specification of functional form and so risks inappropriate specification. Another potential problem is that all deviations are attributed to the inefficiency in regression models, while random shocks might also account for some deviation in the real world. Thus, it is useful to consider how other methodologies could be used to construct alternative performance evaluation systems.

3. Data Envelopment Analysis and Stochastic Frontier Analysis

Charnes, Cooper and Rhodes (1978) proposed a non-parametric relative efficiency measure (referred to hereafter as CCR) based on linking inputs to outputs via an efficiency frontier: DEA. The current study utilizes DEA as well as parametric SFA to develop sets of performance rankings for Peruvian water utilities. The analytics of each approach is summarized in a Technical Appendix available from the authors upon request.[5]

The fundamental problem is how to evaluate a decision-making unit's performance when a system has multiple inputs and outputs. Single-measure gap analysis is often used for performance evaluation and benchmarking. For example, regulators in the water industry commonly use efficiency indicators, such as number of workers per connection and number of connections per 100 households, to assess utilities' performance. However, these measures are not good substitutes for efficiency frontiers, which recognize the complex nature of interactions between inputs and outputs. DEA does not require any specification of the functional form of the production relationship, but develops a frontier relating inputs to outputs.

SFA, on the other hand, constructs the efficiency frontier with a sophisticated economic specification of the production relationship. Its advantage is that the approach attempts to account for the effects of noise in the data. With both DEA and regression analysis, all deviation is attributed to inefficiency. The limitation of SFA is that, in accounting for noise, it assumes that random shock and inefficiency display a specific distribution[6].

---

[5] Books by Cooper, Seiford, and Zhu (2004) and Charnes, Cooper, Lewin and Seiford (1994) survey DEA applications. The book by Kumbhakar and Lovell (2003) provides a comprehensive overview of various SFA methodologies.

[6] With a sufficiently large number of periods, a SFA model using panel data can help in mitigating the strong distributional assumptions that are necessary to disentangle the effects of inefficiency and random noise (Coelli et al.

4. Model Specification and Data

4.1 DEA Model Specification

We are now in a position to analyze the actual performance of 44 EPS from 1996 to 1998. We exclude three companies (company 1, 4, and 5) because of missing data. As Estache et al. (2004) note in their study of South American electric utilities, there are several possible ways to deal with the panel data within the context of DEA. One is to compute a frontier for each period (three cross-sectional analyses in this case) and compare these cross-sectional runs. In this way, one constructs a frontier in each year and can calculate the efficiency of each firm relative to the frontier in each period. Another possibility is to treat the panel as a single cross-section (each firm in each period being considered as an independent observation) and pool the observations. This way, a single frontier is computed, and the relative efficiency of each firm in each period is calculated by reference to this single frontier. In this study, the DEA models are based on pooled data.

The models draw from the earlier research on the characteristics of Peru's water industry (high water loss, low water quality, and excess staff). The inputs of the models are operating costs, water loss, and the number of water connections. The outputs are volume of water billed, the number of customers, coverage of service, and continuity of service. *Operating cost* is calculated by adding direct production costs, sales expenses, administrative expenses, depreciation and provision for uncollectible bills.

Either the network length or the number of water connections can be used to measure the capital of the companies. Because data are missing for network length in 1996-97, we use the number of connections as an indicator of capital. *Volume of water billed* and *the number of customers* are two widely used outputs. According to Corton (2003), Peru has a serious problem with water loss. More than half the water is unaccounted for in more than a third of the companies. Therefore, we use water billed, not water delivered, as an output. We add two additional outputs: *coverage of service* and *continuity of service.* Coverage of service is the second area of efficiency in the SUNASS evaluation system. Continuity of service is an indicator in two SUNASS categories: quality of service (area 1) and management efficiency (area 3). Model 1 is based on the CCR DEA model. Model 2 is based on the Banker, Charnes, and Cooper (BCC) model which drops the constant returns to scale restriction and accommodates variable returns to scale. One limitation of the DEA model is that as the number of variables in a DEA model increases, the discrimination power of DEA models decreases: more firms tend to be on the efficiency frontier. An individual firm is likely to be using relatively less of a particular input or producing more of a particular output, which places it near or on the multidimensional frontier (Rossi and Ruzzier (2000). This problem is more common under the variable returns to scale (VRS) environment since the VRS efficiency scores are relatively higher than CRS efficiency scores. In order to deal with this problem, an input-oriented Super-efficiency VRS model (model 3) is employed.[7] The input-oriented DEA models are used because

2005) and reduce the bias due to unobserved heterogeneity (Farsi et al. 2005). Given the limited number of periods, the panel structure is not explored in this study. A comprehensive review and application of different panel SFA models can be found in Farsi et al. (2005).

[7] The detail descriptions of the models are presented in the Technical Appendix available from the authors.

water utilities are supposed to meet demand (the output is exogenous). Therefore, input quantities appear to be the primary decision variables related to firm efficiency (Coelli et al. 2005). The operating costs have been adjusted by using the GDP Deflator. Table 1 lists the features of each model. The summary statistics of the variables for 1996-98 are shown in Table 2.

Table 1: DEA Model Specification

	Model 1 CCR; Model 2 BCC; Model 3 Super-efficiency
Inputs	Operating costs
	Number of employees
	Number of water connections
Outputs	Volume of water billed
	Number of customers
	Coverage of service
	Continuity of service

Table 2 : Sample Summary Statistics				
Variable	Mean	Standard Deviation	Minimum	Maximum
Physical Outputs				
Water Billed (m3)	5860856	8097041	224870	3.30E+07
Number of customers	98913.44	141488.4	6660	706848
Quality Outputs				
Service Coverage	0.724	0.157	0.21	0.99
Service Continuity	14.343	5.886	1	24
Inputs				
Operation Cost (S/.)	5623718	9136168	114406	4.73E+07
Number of Connections	19781.99	28347.78	1480	147260
Number of workers	121.86	162.97	7	856

Before presenting the efficiency values and rankings obtained by using these two specifications of the DEA model, we summarize the SFA approach.

4.2 SFA Model Specification Augmented by an Input Distance Function

A disadvantage of the production frontier SFA is that it cannot consider multiple inputs and outputs simultaneously.[8] To solve this problem and obtain a more comprehensive ranking, an input distance function was utilized: it is capable of describing the multiple-input and multiple-output production technology of water companies and does not require assumptions about the behavior of the companies (e.g., cost minimization or profit maximization).

Input Distance Function: Following Coelli and Perelman (1999, 2000), Carrington et al. (2002) and Estache et al. (2004), the translog input distance function is adopted, and written as:

$$\ln D_{it} = \alpha_0 + \sum_{m=1}^{M} \alpha_m \ln y_{mit} + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M} \alpha_{mn} \ln y_{mit} \ln y_{nlt} + \sum_{k=1}^{K} \beta_k \ln x_{kit} + \frac{1}{2}\sum_{k=1}^{K}\sum_{l=1}^{K} \beta_{kl} \ln x_{kit} \ln x_{lit}$$
$$+ \sum_{k=1}^{K}\sum_{m=1}^{M} \delta_{km} \ln x_{kit} \ln y_{mit} + \sum_{m=1}^{M} \eta_m \ln y_{mit} T + \sum_{k=1}^{K} \kappa_k \ln x_{kit} T + \phi_1 T + \frac{1}{2}\phi_2 T^2 \qquad (2)$$

($i = 1,2,\ldots,N..$). T is a time trend variable used to capture the time varying effect which is common to all the utilities in a specific time period (e.g. technology change, policy change, regulatory mechanism change, or broad patterns of economic growth.

Here D_{it} represents an input distance: $D = \max\{\rho : (x/\rho) \in L(y)\}$,

$L(y) = \{x \in R_{+}^{K} : xcan.produce.y\}$. The input distance function must be symmetric and homogeneous of degree +1 in inputs. The restrictions required for this are as follows:

$$\sum_{k=1}^{K} \beta_k = 1 \qquad (3)$$

$$\sum_{k=1}^{K} \beta_{kl} = 0 \quad (k = 1,2,\ldots K) \quad \text{and} \quad \sum_{k=1}^{K} \beta_{km} = 0 \quad (k = 1,2,\ldots K) \qquad (4)$$

Those required for symmetry are as follows:

$$\alpha_{mn} = \alpha_{nm}, \quad (m,n = 1,2,\ldots,M) \quad \text{and} \quad \beta_{kl} = \beta_{lk}, \ (k,l = 1,2,\ldots,K) \qquad (5)$$

---

[8] We estimated a production frontier model:

$\ln Volume = \beta_0 + \beta_1 \ln OPEX + \beta_2 \ln \text{Worker} + \beta_3 \ln Connection + \beta_4 t + \beta_5 t^2 + v_{it} - u_{it}$ The result shows that inefficiency accounts for 97 percent of the deviation. All the signs of the coefficients are consistent with the expectation. The OPEX and number of connections have positive and significant impact on the water delivered.

Following Carrington et al. (2002) and Estache et al. (2004), the restrictions can be imposed by normalizing the function by one of the inputs. Because of the homogeneity, if we arbitrarily choose one of the inputs, such as the k[th] input, we can obtain:

$$D\ (x/x_K, y) = D\ (x, y)/x_K \tag{6}$$

$$\begin{aligned}\ln(D_{it}/x_{Kit}) &= \alpha_0 + \sum_{m=1}^{M}\alpha_m \ln y_{mit} + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M}\alpha_{mn}\ln y_{mit}\ln y_{nlt} + \sum_{k=1}^{K-1}\beta_k \ln x^*_{kit} + \frac{1}{2}\sum_{k=1}^{K-1}\sum_{l=1}^{K-1}\beta_{kl}\ln x^*_{kit}\ln x^*_{kit} \\ &+ \sum_{k=1}^{K-1}\sum_{m=1}^{M}\delta_{km}\ln x^*_{kti}\ln y_{mit} + \sum_{m=1}^{M}\eta_m \ln y_{mit}T + \sum_{k=1}^{K}\kappa_k \ln x_{kit}T + \phi_1 T + \frac{1}{2}\phi_2 T^2\end{aligned} \tag{7}$$

where $\quad x^*_k = x_k / x_K$

This can be written as:

$$\begin{aligned}-\ln(x_{Kit}) &= \alpha_0 + \sum_{m=1}^{M}\alpha_m \ln y_{mit} + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M}\alpha_{mn}\ln y_{mit}\ln y_{nlt} + \sum_{k=1}^{K-1}\beta_k \ln x^*_{kit} + \frac{1}{2}\sum_{k=1}^{K-1}\sum_{l=1}^{K-1}\beta_{kl}\ln x^*_{kti}\ln x^*_{kit} \\ &+ \sum_{k=1}^{K-1}\sum_{m=1}^{M}\delta_{km}\ln x^*_{kit}\ln y_{mit} + \sum_{m=1}^{M}\eta_m \ln y_{mit}T + \sum_{k=1}^{K}\kappa_k \ln x_{kit}T + \phi_1 T + \frac{1}{2}\phi_2 T^2 - \ln(D_{it})\end{aligned} \tag{8}$$

($i = 1,2,\ldots,N.$ t =1,…,T**)** and $\quad x^*_k = x_k / x_K$

The distance term $-\ln(D_{it})$ can be viewed as an error term that explains the difference between the observed data points and those points predicted by the estimated transformation function. In the SFA distance function model, the distance term can be replaced with a composed error term, $v_{it} - u_{it}$, to estimate the function as a standard stochastic frontier function. In contrast, Coelli and Perelman (1999, 2000) employ a traditional estimation of distance functions (without two-components errors) and rely on corrected ordinary least squares (COLS). The main difference between COLS and SFA is that COLS attributes all the deviations to inefficiency while SFA models attribute part of the deviations to inefficiency and part of the deviations to random noises. In other words, the SFA models take both inefficiency and random noise into account. We use the software STATA 8.0 and Frontier 4.1 (developed by Tim Coelli) in the estimation. As would be expected, both sets of software yield the same results.

Although the translog function is a flexible functional form, it often violates assumptions required in the input distance function, such as monotonicity or concavity. Furthermore, translog functions can create a multicollinearity problem, which influences the statistical significance of the model (Coelli et al. 2005).[9]

---

[9] We did run the translog model in equation (9). The monotonicity requirement cannot be satisfied. Two out of the six first order terms have wrong signs for coefficients. One of the two is statistically significant at p=0.05. None of the rest first order terms are statistically significant. We also found a very serious multicollinearity problem in the translog function. Many correlations between the regressors exceed 0.85 or even 0.95. In the present study, these problems may prevent us from obtaining meaningful results from the translog specification.

More importantly, given our sample size, the translog functional form will consume too many degrees of freedom (35 regressors in a translog model). Therefore, we employ a log-linear function as an alternative, where the distance function model can be simplified as:

$$-\ln(x_{Kit}) = \alpha_0 + \sum_{m=1}^{M} \alpha_m \ln y_{mit} + \sum_{k=1}^{K-1} \ln x_{kit}^* + \phi_1 t + \frac{1}{2}\phi_2 t^2 - u_{it} + v_{it} \qquad (9)$$

The dependent variable is negative logconnection. Both the truncated normal and half normal model are employed. Since they yield quite similar results, the more general truncated normal model is presented here. The definitions and differences between the two models are described in Kumbhakar and Lovell, 2003. The models are tested using the panel data from 1996 to 1998. The inputs are number of employees, operation cost and number of water connections. The outputs are volume of water billed, number of customers, coverage ratio, and continuity of service. The empirical results indicate that the input distance function SFA model is well-behaved. All the terms have correct signs: the output coefficients are negative, and the inputs coefficients are positive. After controlling for the other inputs, an increase in output should require an increase in the specific input. Notice that the dependent variable is negative input. Therefore, the coefficients of outputs are negative. Similarly, given the outputs, an increase in the other inputs will result in a decrease in the specific input. Therefore, the coefficients of outputs are positive. The impacts of volume billed, number of customers, service coverage and operational costs are statistically significant. The Maximum likelihood estimated parameter "gamma" is 0.732 ($\gamma = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}$), indicating that inefficiency is the main resource of the deviation.[10] See Table 3.

[10] Caudill et al. (1995) suggests that size-related heteroskedasticity of inefficiency variable could lead to biased estimates if the inefficiency term embodies factors "under firm control." We did try the model proposed by Caudill et al. (1995) where variance of inefficiency term is modelled as a function of firm size (log revenue of the firm in a specific year). The estimation results are consistent with the results presented in Table 3 in terms of sign and statistical significance. The mean efficiency scores of the two models are roughly the same (0.89 vs. 0.9) and individual scores are significantly correlated with each other (the correlation is 0.8). Furthermore, the inclusion of the heteroskedasticity adjusted SFA model does not change our interpretation and conclusion qualitatively in the following comparisons.

Table 3: SFA Distance Function Model

Stoc.Frontier Truncated Normal Model					obs =	129
					Wald chi2(8)=	16131.74
Log-Likelihood	=154.02				Prob> chi2=	0.000

neglogconnection	Coef.	Std. Error	z	P>|z|	[95% Conf. Interval]	
logvolumebill	-0.077	0.034	-2.250	0.024	-0.143	-0.010
logcustomer	-0.885	0.035	-25.280	0.000	-0.953	-0.816
logcontinuity	-0.017	0.012	-1.400	0.163	-0.041	0.007
logcov	-0.130	0.029	-4.510	0.000	-0.186	-0.073
logadjustopex	0.063	0.026	2.450	0.014	0.013	0.113
logadjustlabor	-0.011	0.030	-0.370	0.709	-0.071	0.048
t	0.008	0.029	0.280	0.779	-0.048	0.064
t2	-0.003	0.014	-0.250	0.800	-0.030	0.023
_cons	1.772	0.248	7.160	0.000	1.287	2.257
sigma2	0.007	0.003				
gamma	0.732	0.284				
sigma_u2	0.005	0.004				
sigma_v2	0.002	0.002				

5. Data Analysis and Comparison

As noted above, this study focuses mainly on the results for the year 1998 to allow comparison with the SUNASS ranking and Corton (2003). The efficiency scores indicate the presence and extent of inefficiency of input use. For instance, an efficiency score of 0.80 means that the company could reduce its inputs by about 20 percent, while still maintaining its output level. We also find that the SFA techniques yield lower efficiency scores than does DEA. According to Carrington et al. (2002), this difference arises because the estimated frontier binds the data less tightly than with DEA. Fewer efficient decision-making units are identified when the frontier is estimated by SFA analysis. SFA recognizes that some of the distance from the frontier is attributable to random shocks or statistical noise in the data, so it is common to have no efficient decision-making units in a sample. So now we turn to comparisons among the different methodologies. Table 4 displays the summary statistics of the efficiency scores from the different models.[11]

[11] For the input oriented super efficient model, the efficiency scores range from 0 to ∞. The higher the efficiency scores, the more efficient the firms would be. However, the efficiency scores can only give a ranking order of the

Table 4. Summary Statistics of Efficiency Scores from Different Models

Models	Mean	Std.	Min	Max
CCR	0.949	0.066	0.738	1.000
BCC	0.910	0.075	0.732	1.000
Super	1.064	0.540	0.738	6.104
Distance	0.904	0.044	0.797	0.980

Bauer et al. (1998) proposed several sensitivity tests that, if satisfied, would avoid the choice among different technical approaches. In our study, we check the sensitivity in efficiency scores, the rankings, and identification of the best and worst performance. In the first-level test, we compare only the DEA model 1, 2 and SFA Distance Function model because the regression model, the SUNASS evaluation system and Super-efficiency model give the scores in a different scale (not in $[0,1]$). The correlation matrix of the efficiency scores of four models are presented in Table 5.

Table 5: Correlation Matrix of Efficiency Scores

		BCC	CCR	DISTANCE
BCC	Pearson Correlation	1.000	0.736**	0.513**
	Sig. (2-tailed)	.	0.000	0.000
	N	123	123	120
CCR	Pearson Correlation	0.736**	1.000	0.707**
	Sig. (2-tailed)	0.000	.	0.000
	N	123	123	120
DISTANCE	Pearson Correlation	0.513**	0.707**	1.000
	Sig. (2-tailed)	0.000	0.000	.
	N	120	120	129
**	Correlation is significant at the 0.01 level (2-tailed).			

The correlation between BCC and CCR Modes is 0.736, which is significant at the 0.01 level. The correlation between BCC and the SFA distance model is 0.513, and 0.707 between CCR and the SFA distance model, both significant at the 0.01 level. We see that the three parametric and non-parametric comprehensive evaluation models with multiple inputs and outputs yield consistent results, providing a reasonable basis for regulators to design fair scorecards and relevant policies.

---

firms. They are not quantitatively comparable as the CCR and BCC models, because the efficient DMUs are not compared to the same "standards". (The frontier constructed from the remaining DMUs changes for each efficient DMU under evaluation.) (Cooper et al. 2004)

To compare the DEA, SFA production and SFA distance models to the SUNASS evaluation system and Corton's regression model, we test consistency in rankings. Table 6 shows the correlation matrix of the rankings and the rankings provided by different methods.

Table 6: Spearman's Ranking Correlation Matrix of Different Techniques

		SUNASS	REGRESS	BCC	CCR	SUPER	DIST
SUNASS	Correlation Coefficient	1.000	-0.428**	-0.110	-0.396**	-0.121	-0.255
	Sig. (2-tailed)	.	0.007	0.505	0.013	0.481	0.117
REGRESS	Correlation Coefficient	-0.428**	1.000	0.201	0.445**	0.319**	0.117
	Sig. (2-tailed)	0.007	.	0.220	0.005	0.058	0.480
BCC	Correlation Coefficient	-0.110	0.201	1.000	0.643**	0.946**	0.443**
	Sig. (2-tailed)	0.505	0.220	.	0.000	0.000	0.004
CCR	Correlation Coefficient	-0.396**	0.445**	0.643**	1.000	0.731**	0.696**
	Sig. (2-tailed)	0.013	0.005	0.000	.	0.000	0.000
SUPER	Correlation Coefficient	-0.121	0.319**	0.946**	0.731**	1.000	0.416**
	Sig. (2-tailed)	0.481	0.058	0.000	0.000	.	0.009
DIST	Correlation Coefficient	-0.255	0.117	0.443**	0.696**	0.416**	1.000
	Sig. (2-tailed)	0.117	0.480	0.004	0.000	0.009	.

** Correlation is significant at the .01 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

The comparison of the ranking correlations (Table 6) with the current SUNASS evaluation system shows that rankings provided by CCR and Corton's regression model have a negative and statistically significant correlation with the SUNASS rankings. However, if we take the economies scale into account and allow for variable return to scale, the correlation between SUNASS ranking and the three models (BCC, Super-efficiency and Distance function) is not statistically distinguishable from zero. As we mentioned above, there are several potential reasons for this inconsistency. First, the SUNASS ranking system assigns an equal weight to each of the nine sub-dimensions. This method is somewhat arbitrary and may be easy for the firms to manipulate the ranking through unaudited data filings. Firms can choose to improve the dimension that requires minimal outlays and degrade the dimension which is more costly. In contrast, the weights are not pre-assigned in the DEA models or SFA models. They are generated though the linear programming optimization or maximum likelihood estimation. Second, most indicators in the SUNASS scheme lack input-output causative relationships. Only the managerial efficiency category considers the cost issue. This approach is inconsistent with the basic definition of productivity. Therefore, the benchmarking scheme in Peru is basically output-oriented and does not incorporate inputs in a comprehensive manner.

To illustrate this point more clearly, we first calculate six basic output/input ratios for the year 1998 sample: volume billed/OPEX, customer number/OPEX, volume billed/ labor, customer number/labor, volume billed/connections, customer number/connections. We then pick up several points which have the biggest ranking gap between the SUNASS and other techniques and compare these points to the sample[12]. Consider Case 1 (EMAPA PASCO S.A.): its rankings under SUNASS, Corton's regression, CCR, BCC, Super, Distance function are 44, 1, 1, 1, 6, 14 respectively. At first glance, we can find that its service coverage is 59 percent, and its service continuity is only 3 hours; both numbers are substantially lower than the sample mean (75.9 percent, 14.2 hours). These are part of the reason for its lowest ranking in the SUNASS scheme. However, if we dig a little bit deeper and check its six basic output/input ratio, we would find that it has the highest volume billed/OPEX (3.42) and customer number/OPEX (0.065) ratios (the sample means are 1.317, 0.023 respectively). Its other ratios are close to the sample mean. This pattern explains why the firm is on the efficiency frontier in the DEA models. In this case, the Super-efficiency model and distance function model seem to incorporate the two aspects better—yielding a reasonable ranking.

Case 2 provides another example: firm EMAPA COP. Its rankings under SUNASS, CCR, BCC, Super, Distance function are 41, 1, 1, 9, 4 respectively. Its service coverage is 47 percent, and its service continuity is 11 hours, which partially explain its low ranking under the SUNASS scheme. However, if we check its six basic output/input ratios, we would find that it has the highest volume billed/connection and customer number/connection ratios which places it on (or close to) the efficiency frontiers under the DEA and SFA models. Case 3 is firm SEDACAJ S.A., whose rankings under SUNASS, CCR, BCC, Super, Distance function are 2, 37, 37, 34, 33 respectively. Its service coverage is 85 percent, and its service continuity is 20 hours; both of which are significantly higher than the sample mean. These partially explain its high ranking under the SUNASS scheme. If we check its six basic output ratios, its Volume billed/OPEX, Volume billed/labor, Volume billed/connection, customer number/OPEX, customer number/labor and customer number/connection are 0.832, 33965, 330.51, 0.0114, 467.58 and 4.55 respectively, while the sample mean of these six ratios are 1.31, 49430, 287.6, 0.0235, 850.2 and 4.86 respectively. Thus, five indicators are significantly lower than the sample mean. Therefore, its rankings in the frontier models are lower than the median. The above examples show that lack of input-output relation in the SUNASS scheme is one of the main reasons for the inconsistency. In addition, the current SUNASS scheme does not take some other factors such as economies of scale into account.

The result of the regression cost function model proposed by Corton (2003) is somewhat correlated to the DEA models but not correlated to the distance function, which shows the importance of choosing an appropriate model and considering multivariate dimensions. Consistent to what we found in Table 6, the comprehensive parametric and non-parametric evaluation models are statistically significantly correlated to each other. The correlation between BCC and CCR models is 0.643, which is significant at the 0.01 level. The correlation between BCC and Super-efficiency model is 0.946, and the correlation between CCR and Super-efficiency model is 0.731, both significant at the 0.01 level. The stochastic frontier distance function model is positively correlated to all the three DEA models, all significant at p=0.01 level. The correlation between distance function and BCC, CCR, and Super-efficiency models are

[12] This methodology is somewhat inaccurate in the sense of considering the multiple inputs and outputs simultaneously and drawing the whole picture. However, it provides an intuitive explanation regarding the source of the difference between the SUNASS approach and frontier techniques.

0.443, 0.696, 0.416 respectively, which indicates a moderate consistency between parametric and non-parametric models.[13]

If consistency in overall rankings is not met, a third level of consistency involves identifying best and worst performers. As Estache et al. (2004) point out for implementing regulatory policy it is more important to identify the rough ordering of firms rather than obtaining precise measures of efficiency. We compare the results of different techniques in terms of how they identify the best and the worst quartile (ten firms in each). The results are shown in Table 7. The upper triangle contains the fraction of firms classified in the upper quartile for each pair of techniques. The lower triangle shows the same for the lower quartile. The 1998 ranking is used for comparison.

Table 7: Consistency in Identifying Best and Worst Performers
(Upper and Lower Triangles, respectively)

Approach	SUNASS	Regression	BCC	CCR	Super	SFA Distance
SUNASS	1.00	0.00	0.50	0.20	0.40	0.10
Regression	0.10	1.00	0.60	0.40	0.40	0.30
BCC	0.20	0.40	1.00	1.00	1.00	0.60
CCR	0.10	0.40	0.70	1.00	0.60	0.50
Super	0.20	0.30	1.00	0.70	1.00	0.40
SFA Distance	0.30	0.30	0.50	0.80	0.50	1.00

The results show that there is little consistency (10-30%) between the SUNASS ranking and DEA analysis and SFA methods in identifying the worst companies and a little bit more consistency in identifying the best firms (10%-50%). Corton's regression model has some consistency (30%-40%) with the comprehensive frontier models (the DEA models and the SFA distance function model) in identifying the worst companies, and it has more consistency (30-60%) with the comprehensive models (the DEA models and the SFA distance function model) in identifying the best companies. The consistency between the comprehensive evaluation models is higher. The SFA distance function model has 50-80 percent consistency with the DEA models in identifying the worst performance companies and 40-60 percent in identifying the best. The consistency among the three DEA models in identifying the best and worst companies ranges from 70-100 percent.

As Estache et al. (2004) note, the advantage of knowing if the different approaches are consistent in identifying "best" or "worst" firms is that, even if the ranking consistency test fails, a "mild" form of

[13] Since there are non-negligible discrepancies across different measures of efficiency, a comprehensive measure might be useful to regulators in the benchmarking studies. Coelli and Perelman (1999) have suggested combining the results from alternative modeling exercises by using the geometric means of the performance scores for each data point in order to mitigate potential bias of specific methods. This idea is borrowed from the time-series forecasting literature where many authors assert that the mean of the predictions from various models will often outperform any one particular predictive model.

benchmarking regulation can be relied on. This is similar to what the water regulator for England and Wales does when OFWAT publishes the efficiency rankings in the media to increase public pressure on the regulated companies.

Regulators can have some confidence in company classifications that involve similar rankings under different models. In addition, they will need to pay more attention to those with divergent rankings under the various models (illustrated by the three examples we mentioned in explaining the difference between SUNASS and frontier rankings.)

6. Conclusions

Since efficiency evaluation plays such an important role in incentive regulation, analysts and regulators need to exercise great care when selecting ranking techniques from among parametric and non-parametric evaluation models. First, regulatory objectives need to be prioritized, to determine what policymakers view as important. Regulators might want to focus only on cost minimization. In this case, they can choose from among the regression model, COLS (corrected ordinary least squares), SFA production (cost) function model, or single-output DEA models. If regulators want to incorporate other outputs into the analysis, including quality of service, they can choose to use a synthetic evaluation system (like the one utilized by SUNASS), DEA models or SFA input (output) distance function models. The present study shows that the consistency of the efficiency measurement within a specific group of models is high in the case of Peruvian water utilities.[14]

Second, analysts need to be aware of the advantages and limitations of different models in order to select the most appropriate ones for conducting benchmarking studies. As noted, SUNASS assigned an equal weight of 1 to all nine indicators: an output oriented approach that does not consider inputs in a comprehensive way. That commission's logical next step might be to formally prioritize the multiple dimensions of performance, so that rankings better target those areas of greatest importance.[15] The regression model focuses on the cost efficiency of a company, but fails to consider other important factors such as quality of service and coverage of service. Nor does it consider the effect of random shock or statistical noise. In addition, an inherent problem of regression analysis is that it requires specification of functional form, which risks fitting in an inappropriate function. Furthermore, simple regression analysis is limited to only one dependent variable, which might not depict the real world in a sophisticated way.

---

[14] The policy implications of previous studies that have used only one methodology to evaluate relative efficiency would be strengthened by utilizing the types of sensitivity tests described here. Recent studies of other sectors include European banking (Casu and Molyneux, 2003), museums (Bishop and Brand, 2003), Korean container terminals (Cullinane and Song, 2003), Swiss hospitals (Steinmann and Zweifl, 2003), and the Greek agricultural sector (Rezitis, Tsiboukas, and Tsoukalas, 2002).

[15] Lynch, Buzas, and Berg (1994) use hierarchical conjoint analysis to derive weights for dimensions of telephone service quality. The methodology is applicable to water as well—for example, by giving different weights to system expansion and to improvements in service continuity.

DEA models do not have these particular limitations. DEA does not require the specification of a functional form to be fitted, nor does it need to impose weight to the factors. DEA allows for multiple outputs and inputs. In addition, DEA analysis can give us more information than the ranking. It can also be used to evaluate returns to scale and can set the goal for inefficient companies regarding how much they should improve to get on the efficient frontier. However, DEA models have their own imperfections.

The outcome of DEA analysis is sensitive to the selection of the variables for inputs and outputs and to the application of different DEA methods. DEA has been developed in a non-statistical framework, so hypothesis testing is problematic (though bootstrapping techniques have been applied to DEA). In addition, DEA does not account for possible noise. SFA is arguably a better method. It accounts for the effect of the random shocks and statistical noise and can accommodate multiple inputs and outputs by using the distance function. However, it too has potential problems; in particular, the standard SFA method uses a specific assumption on the residual skewness to separate inefficiency from measurement errors. Additional techniques are also available, such as TFA (thick frontier approach) and DFA (distribution-free approach) (Berger and Humphrey 1991; Berger 1993). All these methodologies have their advantages and disadvantages.

A third point is that regulators can select two to three appropriate techniques to construct models and then conduct multi-level consistency tests to compare the outcomes of different methods to determine whether serious inconsistencies emerge. If the analyst uses panel data, regulators should also check whether these efficiency measures are consistent over time. If the consistency tests are satisfied, the regulator can choose one of the techniques with some degree of confidence, or the average ranking could be used. If the tests do not support consistency, regulators should be very cautious about applying financial rewards or penalties based on rankings, since some utilities are likely to challenge the implementation of incentives as reflecting arbitrary procedures. The results for Peru presented here do not invalidate the performance rankings obtained by the regulator, since those rankings are based on more dimensions of utility performance. However, they point out the potential pitfalls of current SUNASS ranking schemes.

References

Andersen, P., and Petersen N.C. (1993), "A Procedure for Ranking Efficient Units in Data Envelopment Analysis," *Management Science*, 39, 10, 1261-1264

Ashton, J. K. (2000), "Cost Efficiency in the UK's Water and Sewerage Industry," *Applied Economics Letters*, 7, 455-458.

Banker, R,D., Charnes, A., Cooper, W. W. (1984) "Some Models for Estimating Technology and Scale Inefficiency in Data Envelopment Analysis," *Management Science*, 30, 9 1078-1093.

Bhattacharyya, Arunava, Elliott Parker, and Kambiz Raffiee (1994), "An Examination of the Effect of Ownership on the Relative Efficiency of Public and Private Water Utilities," *Land Economics* 70(2): 197-209.

Bauer, P., A. Berger, G. Ferrier, and D. Humpherey (1998), "Consistency Conditions for Regulatory Analysis of Financial Institutions: A Comparison of Frontier Efficiency Methods," *Journal of Economics and Business* 50, 85-114.

Berger, A. N. (1993), " 'Distribution-Free' Estimates of Efficiency in the U.S. Banking Industry and Tests of the Standard Distributional Assumptions," *Journal of Productivity Analysis* 4(3): 261-92.

_____, and D.B.Humphrey (1991) "The Dominance of Inefficiencies Over Scale and Product Mix Economies in Banking," *Journal of Monetary Economics* 28, 117-148.

Bishop, P., and S. Brand (2003). "The efficiency of museums: a stochastic frontier production function approach," *Applied Economic* 35, 1853-1858.

Carrington, Roger, Tim Coelli, and Eric Groom (2002), "International Benchmarking of Monopoly Price Regulation: The Case of Australian Gas Distribution," *Journal of Regulatory Economics* 21(2): 191-216.

Casu B., and Molyneux P. (2003). "A comparative study of efficiency in European banking," *Applied Economics* 35, 1865-1876.

Charnes, A., W.W. Cooper, and E. Rhodes (1978), "Measuring Efficiency of Decision Making Units," *European Journal of Operational Research* 2, 429-44.

_____, W. W. Cooper, A. Y. Lewin, and L. M. Seiford (1994), *Data Envelopment Analysis: Theory, Methodology and Application*. Boston: Kluwer Academic Publishers.

Clarke, George R.G., Katrina Kosec, and Scott Wallsten (2004), "Has Private Participation in Water and Sewerage Improved Coverage? Empirical Evidence from Latin America," Working Paper 04-02 (January), AEI-Brookings Joint Center for Regulatory Studies.

Coelli, T.J., A. Estache, S. Perelman, and L. Trujillo (2003), *A Primer on Efficiency Measurement for Utilities and Transport Regulators*. Washington, D.C.: World Bank Publications.

_____, D. S. Prasada Rao, Christopher J. O'Donnell, and G. E. Battese, (2005), *An Introduction to Efficiency and Productivity Analysis*. Second Edition. Boston: Kluwer Academic Publishers.

_____, and S. Perelman (1999), "Theory and Methodology: A Comparison of Parametric and Non-parametric Distance Functions: With Application to European Railways," *European Journal of Operational Research* 117: 326-39.

_____, and______. (2000) Technical Efficiency of European Railways: a Distance Function Approach, *Applied Economics*, 32, 1967-1976.

Cooper, W.W., L.M. Seiford, and K. Tone (2000), *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver*. Boston: Kluwer Academic Publishers.

_____, L.M. Seiford, and J. Zhu (2004), *Handbook on Data Envelopment Analysis.* Boston: Kluwer Academic Publishers.

Corton, Maria Luisa (2003), "Benchmarking in the Latin American Water Sector: The Case of Peru, *Utilities Policy* 11: 133-42.

Cubbin, John, and George Tzanidakis (1998), "Regression versus Data Envelopment Analysis for Efficiency Measurement: An Application to the England and Wales Regulated Water Industry," *Utilities Policy* 7(2): 75-85.

Cullinane, K., and D-W. Song (2003). "A stochastic frontier model of the productive efficiency of Korean container terminals," *Applied Economics*,35, 251-267.

Estache, Antonio, and Martin A. Rossi (2002), "How Different Is the Efficiency of Public and Private Water Companies in Asia?" *World Bank Economic Review* 16(1): 39-148.

_____, and Eugene Kouassi (2002), "Sector Organization, Governance, and the Inefficiency of African Water Utilities," World Bank Policy Research Working Paper 2890.

_____, Martin A. Rossi and Christian A. Ruzzier (2004), "The Case for International Coordination of Electricity Regulation: Evidence from the Measurement of Efficiency in South America," *Journal of Regulatory Economics* 25(3): 271-95.

Farsi, M., Filippini, M., and M. Kuenzle, M. (2005). "Unobserved heterogeneity in stochastic cost frontier models: an application to Swiss nursing homes," *Applied Economics* 37, 2127-214.

Fiordelisi, F., and P. Molyneux (2004). "Efficiency in the factoring industry," *Applied Economics* 36, 947-959.

Hunt, L.C., and E. L. Lynk (1995), "Privatization and Efficiency in the UK Water Industry: An Empirical Analysis," *Oxford Bulletin of Economics and Statistics* 57: 371-88.

Jondrow, J., C.A.K. Lovell, I.S. Materov, and P. Schmidt (1982), "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model," *Journal of Econometrics* 19(2/3): 233-38.

Kirkpatrick, C., D. Parker, and Y-F Zhang (2004), "State versus Prive Sector Provision of Water Services in Africa: A Statistical, DEA and Stochastic Cost Frontier Analysis," Working Paper no. 70, Center on Regulation and Competition, University of Manchester, UK.

Latruffe, L., Balcombe, K., Davidova, S., and Zawalinska (2004). "Determinants of technical efficiency of crop and livestock farms in Poland," *Applied Economics* 36, 1255-1263.

Lin, Chen (2005), “Service Quality and Prospects for Benchmarking: Evidence from the Peru Water Sector” *Utilities Policy*, 13. 230-239.

Lynch, John, Tom Buzas, and Sanford Berg (1994), "Regulatory Measurement and Evaluation of Telephone Service Quality," *Management Science* 40(2): 169-94.

Rezitis, A., K. Tsiboukas, and S. Tsoukalas (2002). “Measuring technical efficiency in the Greek agricultural sector,” *Applied Economics* 34, 1345-1357.

Saal, D.S. and Parker, D., 2001. Productivity and Price Performance in the Privatised Water and Sewerage Companies of England and Wales. *Journal of Regulatory Economics* 20(1): 61-90.

Smith, P. (1990), “The Use of Performance Indicators in the Public Sector,” *Journal of the Royal Statistical Society, Series A* 153**,** 53–72.

Steinmann, L., and P. Zweifel (2003). “On the (in)efficiency of Swiss hospitals,” *Applied Economics* 35, 361-370.

Torero, Maximo, and Alberto Pasco-Font (2003), “The Social Impact of Privatization and the Regulation of Utilities in Peru,” in *Utility Privatization and Regulation: A Fair Deal for Consumers?* Cecilia Ugaz and Catherine Waddams Price, eds. Cheltenham, UK: Edward Elgar Publishers, 257-88.

Tupper, Henrique Cesar, and Marcelo Resende (2004), “Efficiency and Regulatory Issues in the Brazilian Water and Sewage Sector: An Empirical Study,” *Utilities Policy* 12: 29-40.

Wallsten, Scott, and Katrina Kosec, 2005. “Public or Private Drinking Water? The Effects of Ownership and Benchmark Competition on U.S. Water System Regulatory Compliance and Household Water Expenditures.” AEI-Brookings Joint Center Working Paper 05-05.