Econometric Measures of the Relative Efficiency of Water and Sewerage Utilities in Brazil.

Guillermo Sabbioni*

October 5, 2006

This study uses econometric techniques to measure the relative performance of water and sewerage utilities in Brazil in 2002. After a brief review of possible methodologies and an explanation for the study's reliance on a cost function approach, three alternative specifications of a stochastic cost frontier are utilized to rank Brazilian firms. Results show a significant stability of the efficiency rankings across the three cost frontiers. Such rankings yield scorecards that can help policymakers identify firms that provide value to citizens.

JEL codes: D24, D61

Keywords: Brazil, Efficiency, Stochastic Frontier, Water and Sewerage.

* University of Florida, Warrington College of Business, 324 MAT, PO BOX 117140, Gainesville FL 32611, guillermo.sabbioni@cba.ufl.edu.
I want to specially thank Sanford Berg, whose comments on an earlier version are much appreciated. I also thank Fernando Damonte, Maria Luisa Corton, Chen Lin and Hamilton Silva for their suggestions.

Econometric Measures of the Relative Efficiency of Water and Sewerage Utilities in Brazil.

Early work on the relative performance of water and sewerage (WS) utilities by Crain and Zardkoohi (1978) tried to determine whether private U.S. water utilities attained a more efficient level of operation than public ones. Since then, a number of papers have been published on the efficiency of WS utilities. Some authors have also focused on the private vs. public issue[1], while others have tested other hypotheses, like the existence of economies of scale, economies of scope, or the possible homogeneity or homotheticity in the production technology.[2] Data availability limited the types of studies: early papers focused mostly on utilities in the U.K. and U.S. because these countries pioneered the collection and publication of data on WS firms.[3]

Until a decade ago, little research was conducted into the efficiency of WS utilities in developing countries. Then, studies began to address the performance of water systems using quantitative techniques. A number of papers focused on Asian and African water utilities, most of them supported by the World Bank.[4] These studies provided insights for countries implementing World Bank policies to increase coverage and quality of WS services in their regions.

Recently, two papers on Latin America have added additional evidence regarding the differential performance of WS utilities. One, a Brazilian study by Tupper and Resende (2004), used a modified data envelopment analysis (DEA) to quantify the relative efficiency of twenty state WS companies for 1996-2000.[5] Their study was the first to relate three inputs (labor expenses, operational costs and other operational costs) to four outputs (water produced, treated sewerage, water population served, and treated

[1] For example, Feigenbaum and Teeples (1983), Estache and Rossi (1999), Estache and Rossi (2002), and Estache and Kouassi (2002).
[2] For example, Fox and Hofler (1986), Kim (1987) and Saal and Parker (2000).
[3] Many studies were prepared under the auspices of OFWAT, the economic regulator for the water and sewerage industry in England and Wales. See www.ofwat.gov.uk.
[4] For example, Estache and Rossi (1999), Estache and Rossi (2002), and Estache and Kouassi (2002).
[5] Another study, by Corton (2003), examined Peruvian WS firms.

sewerage population served). In addition, Tupper and Resende applied more standard econometric techniques to partly explain the relative efficiency of the firms.

The current study utilizes a larger sample of Brazilian firms with data from 2002. Rather than DEA, this study applies stochastic cost frontier models to the expanded set of firms. The results illustrate how another methodology can be utilized to evaluate efficiency of WS utilities. This study does not attempt to compare the strengths and limitations of the parametric and nonparametric approaches to benchmarking WS utilities. Although the issue is an important one, more specific literature has addressed the topic.

The structure of the paper is as follows: Section I presents alternative econometric methodologies for evaluating the relative performance of firms: production and cost functions. The next section describes the cost function approach and the data utilized in the study. Section III presents a more detailed analysis of the variables included in the analysis.[6] The utility rankings are presented in the next section, based on three stochastic cost frontier models. The concluding Section V offers suggestions for future research.

I. Possible methodologies: production and cost functions

Measuring the relative performance of firms is a complicated issue. There are two main methodologies, each with advantages and disadvantages: econometrics and the relatively new non-parametric DEA. It is still not clear which one should be used in each particular situation[7]. Previous studies of WS utilities have generally utilized one or the other. Moreover, some studies have tried to use both methodologies when ranking firms, examining the consistency of the efficiency rankings.[8]

This paper relies on econometric techniques. It is well-known that non-parametric DEA can also answer some important questions. Nevertheless, this study focuses on parametric approaches to estimate

[6] A complete and detailed description is given in **Appendix C**.
[7] Some authors have established the framework that they think would be the best for each methodology; see, for example, Estache and Rossi (1999), and Berg and Lin (2005).
[8] See Cubbin and Tzanidakis (1998) and Berg and Lin (2005).

key parameters and test the relative importance of variables.[9] When choosing to measure efficiency with econometric techniques, analysts have two options: cost functions and production functions. Neither is perfect, given data limitations. The preference for one or the other depends on the particular circumstances in five areas:

1) *Operating environment*: This is an institutional factor encompassing the framework within which the firm performs its activities. Under an assumption of profit maximization, a production function may be the natural choice for a firm that could select its output level. However, if the utility has an "obligation to service", it will be required to produce the output level demanded by customers. The production function approach may not be suitable for this situation.

2) *Possible endogeneity*: A second issue is the endogeneity of input quantities when utilizing a production function. This problem can be addressed by using a system of simultaneous equations, where input demands and the production function are jointly estimated, which complicates the procedure from the econometric point of view.[10] Unlike a production function, a cost function is not likely to present this endogeneity problem, although the underlying structure of production will not be identified.

3) *Data limitations*: Data availability is crucial in the selection of the empirical model. The utilization of a production function requires data on the amount of physical units of all inputs and outputs, while the cost function requires prices of inputs, rather than quantities, for generating the observed outputs. A problem with cost functions is that relevant input prices are rarely reported in databases. Only the price of labor, calculated as the sum of salaries paid divided by the number of people working in the firm, is usually available (and even then, indirectly).[11] Some papers have tried to calculate the price of

[9] For more complete descriptions of the differences between DEA and econometrics in the context of WS utilities, see Cubbin and Tzanidakis (1998) and Berg and Lin (2005).
[10] Another possible solution is the Instrumental Variable (IV) technique, which corrects for the endogeneity of input quantities. See Estache and Kouassi (2002) for an application of this methodology in the WS utilities environment.
[11] Some U.S. studies had data on the level of wages across the different states, and so did not have to calculate the average salary in that way.

capital to replicate a traditional microeconomics textbook technology[12]; however, no convincing methodology has emerged in this area.

4) *Output definition*: An issue when using either the production or the cost function approach is the definition of output. Variables used in past studies have included the number of customers served, the number of connections, and/or the volume of water delivered. All of these are possible measures of the output of WS firms. Some studies consider the possibility of more than one output.[13] This multiplicity is difficult to address in the production function framework, since it constrains us to choose just one output. However, if we decide that a cost function is the proper choice, each output can be included as an exogenous variable in the regression, as long as the firm does not select its level.

5) *Technology specification*: Lastly, in both the cost and the production function approaches, the value of exogenous parameters can be included if we assume that they affect the technology used by the firm. Some of these environmental variables have been included in previous studies: the proportion of water delivered to residential ("small") vs. commercial ("big") customers, the density of population in the area served, the proportion of water coming from ground vs. surface sources, the proportion of connections that are metered, the type of treatment applied to the water (chlorination, desalination, disinfection), the level of corruption and governance (when doing cross-country studies), private vs. public ownership, regional dummies (to capture topological factors), the maximum storage capacity, the proportion of water purchased vs. self-supplied, and the number of districts served, to list a few cost drivers.

Even though most of these variables seem reasonable, many studies that include these factors do not find significant coefficients. The roles of particular variables have been disputed. For example, Feigenbaum and Teeples (1983) argue that density of population has a positive effect on cost[14], while all

[12] See Kim (1987) for an estimation of the price of capital as the long-term interest rate of the debt plus a charge of 2% for amortizations.
[13] See Fox and Hofler (1986), Kim (1987), Saal and Parker (2000) and Saal and Parker (2001).
[14] They say: "We would expect that it is more costly to supply more densely developed service areas, which requires more hydrants, higher water pressure and greater peak capabilities for fire protection," and they confirm that result in their paper (1983, p.674).

other researchers state that this variable should have a negative effect. Thus, determining how exogenous variables affect cost requires a careful model specification.

In summary, addressing the relative performance of WS utilities with econometric techniques is difficult, and there is no agreement on which approach is best. Therefore, this study should be viewed as exploratory: improving our understanding of cost drivers. The results will be suggestive, supporting some studies and raising questions about others.

II. The cost function approach adopted

Having described the main differences between the cost function and the production function approaches, this study utilizes the former as more appropriate, based upon the five issues stated above: characteristics of the operating environment of WS utilities, the ability to deal with multiple outputs, the lack of endogeneity problems, data availability, and technology specification.[15]

We assume that a WS utility can maximize its profits by minimizing the cost of producing some exogenously given output level, subject to the available technology (i.e., the production function). The solution to this optimization (cost minimization) problem is a cost function:

$$C = C(Y, W, Z, D), \qquad (1)$$

where Y is the output vector, W is the vector of input prices, and Z and D are vectors of environmental characteristics. We assume that this cost function $C(\bullet)$ can be decomposed as the product of two functions $f(\bullet)$ and $g(\bullet)$. The function $f(\bullet)$ will have input prices W and outputs Y as arguments. And the function $g(\bullet)$

[15] The majority of relevant papers preferred the cost function approach: Crain and Zardkoohi (1978), Feigenbaum and Teeples (1983), Kim (1987), Teeples and Glyer (1987), Cubbin and Tzanidakis (1998), Saal and Parker (2000), Estache and Rossi (1999), Estache and Rossi (2002), Corton (2003) and Tupper and Resende (2004). However, a number of scholars have used production function approaches: Byrnes et al. (1986), Fox and Hofler (1986) and Estache and Kouassi (2002).

will incorporate the exogenous variables (*Z* and *D*) that will affect the technology of the firm. Our cost function is then:

$$C = f(Y,W)\,g(Z,D) \tag{2}$$

We choose a Cobb-Douglas form for *f(Y,W)*:

$$f(Y,W) = A\prod_{i=1}^{n} Y_i^{\alpha_i} \prod_{j=1}^{m} W_j^{\beta_j} \quad \text{where A, } \alpha_i \text{ and } \beta_j \text{ are all parameters.} \tag{3}$$

And we define *g(Z,D)* in the following way:

$$g(Z,D) = \prod_{k=1}^{s} Z_k^{\gamma_k} e^{\sum_{d=1}^{r} \delta_d D_d} \quad \text{where } \gamma_k \text{ and } \delta_d \text{ are all parameters.} \tag{4}$$

The explicit form of the cost function *C(•)=f(•)g(•)* will then be:

$$C = A\prod_{i=1}^{n} Y_i^{\alpha_i} \prod_{j=1}^{m} W_j^{\beta_j} \prod_{k=1}^{s} Z_k^{\gamma_k} e^{\sum_{d=1}^{r} \delta_d D_d} \tag{5}$$

After applying natural logarithms to both sides, the linear cost function has the following form:

$$\ln C = \ln A + \sum_{i=1}^{n} \alpha_i \ln Y_i + \sum_{j=1}^{m} \beta_j \ln W_j + \sum_{k=1}^{s} \gamma_k \ln Z_k + \sum_{d=1}^{r} \delta_d D_d \tag{6}$$

The logarithmic form of *Y*, *W* and *Z* provides an intuitive interpretation for the coefficients as elasticities. The reason for the separation of *Z* and *D* in *g(•)* is clear: this allows us to have environmental variables in logarithmic form while the dummies are not in logarithmic form. The three alternative specifications presented in this work will differ in the choice of outputs *Y* and the exogenous factors *Z* and *D*. The W vector of input prices remains the same across the models.

<u>Data</u>

Brazil has published data pertaining to the performance of WS utilities since 1995. The data are publicly available on the web page of the National Sanitation Information System (*Sistema Nacional de Informacoes sobre Saneamento, SNIS*). The SNIS program is administered by the Sanitation Modernization Program (*Programa de Modernizacao do Setor Saneamento*) under the National Secretary of Environmental Sanitation (*Secretaria Nacional de Saneamento Ambiental*).

The online data are split into several files, each containing only a certain group of variables (financial, descriptive, operational, etc) and a certain type of firm (regional, micro-regional or local).[16] These spreadsheets were pooled together for this work. The focus on 2002 data provided 280 observations for the cross-sectional analysis. However, missing observations reduce the effective size of the sample, depending on the variables chosen.

As the Visao Geral (2002) notes, the collected data represent a significant proportion of the water sector in Brazil: the WS utilities reporting data serve almost 134 million people with water, which is more than 94 percent of the nation's urban population. Regarding sewerage, the WS utilities reporting data serve around 101 million people, representing 71 percent of Brazil's urban population.

[16] The state firms analyzed by Tupper and Resende (2004) are labeled regional in the present study. This paper expands the sample to include micro-regional and local firms.

III. Three possible specifications for the cost function

Having introduced the mathematical form of the cost function, this section turns to the results from the OLS regressions, where the dependent variable is the annual operating cost of the firm, excluding depreciation charges.[17]

The first results consist of specifications that deliberately exclude environmental variables *Z* and *D*. Although important information is missing, the specifications show that a high percentage of the variation in the operating cost of the firms is explained with only output measures *Y* and input prices *W*. The R^2 for these regressions is higher than 0.90.

This first set of results consists of three different specifications for the cost function, depending on the output variables chosen: Volumes Model, Population Model, and Connections Model. These versions are subsequently labeled as the Basic OLS Models. Each model has one output variable related to the water provision service and one output variable related to the sewerage collection service[18], as follows:

- The Volumes Model includes the following output variables:

- WATER VOL PROD: the (ln of the) amount of water produced in 2002, in thousands of m^3.
- SEWER VOL COLL: the (ln of the) amount of sewerage collected in 2002, in thousands of m^3.

- The Population Model includes the following output variables:

- POP SERV WATER: the (ln of the) number of people served with water services in 2002.

[17] This study does not consider capital expenditures. All investments that logically generate depreciation charges are excluded. While network capacity is important, our evaluation of efficiency focuses only on the annual operating cost of the firms.

[18] This significantly reduced the original size of the sample (280 observations) because only 163 firms are engaged in both activities. So every time a sewerage output variable was included, the size of the sample was dramatically reduced.

- POP SERV SEWER: the (ln of the) number of people served with sewerage services on the same year.

- The Connections Model includes the following output variables:

- WATER CONN: the (ln of the) number of water connections in 2002.
- SEWER CONN: the (ln of the) number of sewerage connections in 2002.

Regarding the W vector of input prices present in all regressions, the variable WAGE was estimated in the usual way, as the ratio of total salaries paid divided by the number of workers in the firm. This variable is hence measured in R$/year.[19] Note that this is the price of the most important input within the operating cost.[20]

Summary statistics of the variables used in the Basic OLS Models are depicted by Table 1 below:[21]

Table 1

We show below the results from the three Basic OLS Models, where we ignore *g(Z,D)*, so *C(•)=f(•)*. Table 2 shows the independent variables used (all in ln terms), their estimated coefficients, and their t-values in parenthesis:[22] [23]

Table 2

---

[19] R$ denotes "Reales", the brazilian currency.
[20] Labor accounts for approximately 44% of operating expenses. The second-most important input is Purchases from Third Parties, which are composed mainly of salaries according to the Visao Geral (2002)). The next most important input is Energy, accounting for 14% of total operating cost.
[21] Missing values create the difference in the number of observations across models.
[22] In all tables, * indicates significance at the 10% level, ** at 5%, and *** at 1%.
[23] Likely heteroskedasticity attributable to the great difference in the size of the utilities is controlled for in all regressions. Thus, the coefficient variances are estimated with the Huber/White format, without the assumption of identically distributed errors. In fact, the presence of heteroskedasticity is confirmed when we eliminate the robust command and the t-ratios dramatically increase because of a great reduction in the errors of the estimated coefficients.

Some preliminary and expected conclusions can be extracted from Table 2. First, we see that the WAGE coefficient is always positive and significant, confirming the hypothesis that the operating cost of the firm will increase if the price of labor goes up. Second, we see that no matter the choice of outputs (Volumes, Population or Connections), all the coefficients are positive and significant in the three regressions. These models confirm the hypothesis that it is costly to produce more output, other things being equal.[24] [25]

Finally, variations of the Basic OLS Models were evaluated to shed light on the environmental issues that may affect the technology of the firms. So the last step was to check the effect of exogenous factors by incorporating *g(Z,D)* in the analysis. Environmental variables were tested one by one to isolate their effect on the operating cost and evaluate them in detail. Additional variables were included only if they added significant explanation to the operating cost by having stable, meaningful and statistically significant coefficients. It is important to remember that having WAGE and just one output variable for both water and sewerage has provided a high R^2 and significant coefficients for the Basic OLS Models, as shown in Table 2.

The procedure used to determine the inclusion of exogenous factors is explained in **Appendix C**, which contains the Final OLS Models. The environmental variables finally incorporated into *g(Z,D)* are the following:

- PURCH WATER is a dummy equal to 1 if the utility purchases water from another utility.
- HHOLD WATER CONS is the (ln of the) average liters of water that a household consumes, in m^3 per month.
- SOUTHEAST is a dummy equal to 1 if the firm is located in the southeastern region of Brazil.

[24] As a robustness check, regressions were run for the Basic OLS Models for the year 2001, with similar results shown in **Appendix A**.

[25] A possible variation of the Basic OLS Models, modified to include two different measures of output for both water and sewerage at the same time (four output variables in total), was explored to see whether adding more output measures enabled us to better explain the variation in operating cost. This alternative specification was discarded because of unsatisfactory results. **Appendix B** shows that analysis.

- NORTHEAST is a dummy variable equal to 1 if the firm is in the northeastern region.
- WATER VOL MET is the (ln of the) volume of water that is metered, in m^3.

Summary statistics of the variables used in the Final OLS Models are depicted below in Table 3:[26]

Table 3

In the Final OLS Models shown in Table 4 below, the cost function now takes the complete form $C = f(Y,W)g(Z,D)$:[27]

Table 4

The stochastic cost frontier versions of these models are estimated in the next section, so they can be used to construct an efficiency ranking of the firms. The intuitive explanations for the signs and significance of the environmental variables just added are provided in **Appendix C**, where the process leading to the inclusion of exogenous vectors *Z* and *D* in the regressions is also explained.

IV. Rankings of the Regional Firms

To estimate the efficiency level of the regional firms, we run three stochastic cost frontier models. OLS regressions were valid when trying to identify the significant variables for most of the variation in operating cost, but the construction of an efficiency ranking of the firms needs to account for both inefficiency and randomness in the error term. Otherwise, we risk labeling the error term of a firm as 100 percent inefficiency when some disturbance in the data is perfectly plausible. Hence, we performed three

[26] Again, missing values create the difference in the number of observations across models.
[27] **Appendix D** shows the Final OLS Models for 2001, with satisfactory results in terms of stability, sign and statistical significance of all the variables.

stochastic cost frontier regressions (using maximum likelihood estimation, MLE) with the variables listed in Table 3.

The error term from these regressions is asymmetric, since it is composed of an error term ε (that accounts for the random noise in the data) and a non-negative term μ (that specifically accounts for the deviation in cost attributable entirely to inefficiency).[28] Both components are assumed to add up to the total deviation between the observed operating cost and the operating cost estimated from the regression.

The results from these MLE optimizations are shown in Table 5 below:

Table 5

Notice the strong stability of the value and sign of the coefficients from the OLS versions (Table 4) and the stochastic cost frontier versions (Table 5). Also observe that all the variables remain highly significant across both tables.

Finally, Table 6 below ranks the regional firms in 2002 according to the value of the inefficiency term μ:[29] [30]

Table 6

The four top firms according to our stochastic cost frontier models for 2002 are:

- CAGECE/CE
- COPASA/MG
- SABESP/SP
- CESAN/ES.

[28] The non-negative error term is assumed to follow a half-normal distribution. Other distributions (like the exponential) produce very similar results.
[29] We use a **bold** font for those firms that appear in the top and bottom five positions according to all three cost frontiers.
[30] Recall that the state firms analyzed by Tupper and Resende (2004) are labeled regional in the present study. This paper also includes micro-regional and local firms in the sample.

The three bottom firms according to the same methodology for 2002 are:

- CASAL/AL
- CAESB/DF
- CAESA/AP.

The consistency across the three models lends confidence to the above classification. The firms in both groups are consistently ranked as being in the top or bottom, regardless of which output variable is chosen. Appropriate exogenous factors were incorporated into the models, controlling for the features that undoubtedly condition the technology available to the firm.

V. Concluding observations

This study used OLS to identify variables affecting operating cost of WS utilities in Brazil for 2002. Then, the model was estimated with a stochastic cost frontier in order to construct an efficiency ranking. A cost function rather than a production function was used because of its strengths in the context of WS utilities. Three alternative versions of a cost function were developed, mainly differing in the output variables chosen. These cost functions incorporated exogenous parameters that affect the available technology of the firm. In constructing an efficiency ranking for the regional WS firms, the results suggest consistency in the position that each regression assigned to each utility. In particular, a few firms were repeatedly ranked in the top or bottom of the list for 2002. This result indicates some robustness in the rankings.

Clearly, much work remains. For the purpose of rewarding good performance and penalizing weak performance, scholars and practitioners need to develop benchmarking procedures that can pass legal challenges. As someone said, “If you torture the data enough, they will confess.” Econometricians must contribute to the debate with a sense of humility and with a pragmatic orientation.

The process must continue to build on the pioneering work of those whose work is cited in the references. We are beginning to obtain results that can be used by those implementing policy. In particular, the publication of league tables is one way to put pressure on the weakest performing WS utilities. Similarly, the managers of WS utilities in the top 20 percent might be awarded some share of the cost savings that can be attributed to their efforts. Those promoting improvements in WS sector performance can take steps to reduce production costs and free up cash flows for network rehabilitation and expansion. Identifying, implementing, and evaluating good incentive systems represent a challenge for regulators.

Appendix A: Basic OLS Models for 2001

Table A-1

We see the stability of the Basic OLS Models from 2001 to 2002, according to the value of R^2, and the significance and sign of all the coefficients. (Compare this table with Table 2.)

Appendix B: Mixed Basic OLS Models

Mixed Basic OLS Models try to capture the multiplicity of outputs by simultaneously including two output variables for water and two output variables for sewerage. There is no substantial increase in R^2, but there are dramatic changes in the value, sign and significance of many output variables:

Table B-1

In spite of the poor general performance, there are two key conclusions from the Mixed Basic OLS Models. First, WAGE remains positive and significant. Unlike in Table 2 for the Basic OLS Models, the coefficient is now more similar across the three specifications. This result occurs since there is not a "pure" Volumes model anymore, which is the one that gave us a different value for the WAGE coefficient in the Basic OLS Models in Table 2.

Second, the Basic OLS Models can be used to derive further results with reasonable confidence about having a strong measure of outputs, and about not missing anything related to their multiplicity. Combining two measures for the same output dramatically changes the results because of a severe colinearity problem, as shown in the next table:

Table B-2

Appendix C: Inclusion of Exogenous Factors

Five analyses addressing issues beyond the control of the firm's managers are reported in this appendix. The results support the inclusion of exogenous factors *Z* and *D* in the regressions.

First, we evaluate the possibility of regional influences on operating costs. It could be that the location of a firm may affect its efficiency, for example through the source of water.[31] Hence, we check the significance of four regional dummy variables in the three Basic OLS Models:

Table C-1

At first glance, there is no evidence that any region affects operating costs in any particular direction (at least, in all the specifications). Furthermore, the R^2 does not increase significantly,

[31] This could be a good substitute for ground vs. surface sources of water as well as energy price, neither of which is available in the database. According to the literature, these could be called regional fixed effects variables. There are a total of five regions, with no dummy included for the west central region.

considering that we are adding four new variables. Only the Population model shows two significant dummy variables at the 5 percent level of significance (Northeast and Southeast). And we see that the sewerage output variable is not significant in either the Population or Connections model.

To ensure that we do not miss important information, each model was put through the process of eliminating the less significant regional variables one by one, to evaluate the consistency of the regional effects among the three specifications. After this procedure was performed, the Basic OLS Models including significant regional dummies are as follows:

Table C-2

As it can be seen, no regional dummy remained in the Volumes model (a replication of the basic Volumes model). It seems that if both volumes of water and sewerage are present, knowing where the utility is located brings no extra information for explaining operating costs. However, Northeast (with a negative coefficient) and Southeast (with a positive one) were kept in both the Population and the Connections models because of significant effects.[32]

The second analysis of exogenous factors addresses the constraint that some utilities face when they do not own enough water to satisfy their demand, and so they must purchase water from other firms. A dummy variable is added to control for this issue and to check for any effect on the operating cost.[33] The variable PURCH. WATER is a dummy equal to 1 if the utility purchases water, 0 otherwise. Theory does not suggest an expected sign for this variable:

Table C-3

[32] It can also be noted that as in the first versions with regional dummies, the sewerage variables in the Population and Connections models are not significant when including the Northeast and Southeast variables. This substitution may deserve more attention.

[33] Regional dummies are excluded in Table C 3, since we analyze each set of environmental factors separately. The variables working well in this way are put together in the final version at the end of this Appendix.

There is some evidence that the need to purchase water from other firms increases operating cost, but the significance of this result is verified only for the Volumes model.

The third set of results concerns the effect of population density on operating cost. We want to check whether having customers more densely located reduces the operating cost of the firm, as is usually argued.[34] WATER CONN P/ KM is the (ln of the) number of connections per kilometer of water network, a measure of the density of water connections. SEWER CONN P/ KM is the (ln of the) number of sewerage connections per kilometer of sewerage network. In principle, both variables have negative expected coefficients. Counter to the prediction, no negative and significant coefficients were found for the density variables in all three specifications. By eliminating insignificant variables one by one, we end up with the following Basic OLS Models including density measures:

Table C-4

Notice that the Volumes model does not accept any of the density variables with significance, so in that case we have the basic model again.[35] But density of water connections contributes to both the Population and Connections models with reasonable statistical significance (10%) and the expected negative sign. Sewerage density, in contrast, proved to be insignificant in all models.

The fourth analysis of environmental factors was the evaluation of whether water metering has a positive impact on operating costs. Although this is not necessarily exogenous to the firm, it could explain differences in operating cost not attributable to inefficiency. Hence, it would be valid to include such a measure in the model. WATER VOL MET is the (ln of the) volume of water that is metered, measured in m^3:

[34] Theory specifically stresses the importance of density for investment costs in fixed assets, which are said to be lower in more densely populated areas. The density of population in the areas served would be the best measure to account for the demographic characteristic, but these data are unavailable.
[35] The result is somewhat similar to the one with the regional dummies; i.e. they do not fit in the Volumes model (see Table C 1).

Table C-5

The results indicate that the volume of water metered has a positive effect on operating costs. The result is maintained across the three specifications, although the value of its coefficient is similar only in the Population and the Connections model.[36]

The last set of regressions with environmental variables tests the hypothesis that bigger customers reduce the operating cost of the firm, addressing issues raised in earlier works regarding residential vs. commercial customers (even though that distinction is not made in our database). The variable AVG HHOLD W CONS is the (ln of the) average liters of water that a household consumes in m^3 per month. This variable is expected to have a negative coefficient since (from the operating cost point of view) it should be cheaper to serve a few large customers than many small ones, keeping total volume of water constant.[37] The results are presented below:

Table C-6

It is evident that only the Volumes model accepts a significant measure of average water consumption, which means that the greater the volume of water delivered to a customer, the lower the total operating cost for the utility.

We finally consolidate all the exogenous variables that worked well independently in the Basic OLS Models to examine their effects on operating cost. The Volumes model remains stable, with all the environmental factors being significant. The Population and Connections models show some insignificant variables but still keeping the same earlier signs. The final step is to eliminate insignificant variables one by one, which yields the three Final OLS Models:[38]

[36] This is similar to the behavior of WAGE in the Basic OLS Models.
[37] Think of the cost of customer service and billing, for example.
[38] We again observe the elimination of the sewerage output measure when including exogenous factors in the Population and Connections models. This increases the size of the sample since firms that supply water only are now included in the regressions.

Table C-7

After having addressed all the environmental factors analyzed in this paper, it is worth mentioning two variables usually included in the previous literature but not incorporated in our models: water losses and length of network. Water losses could certainly explain differences in operating cost across utilities, but that is something that should be handled by the management of the company. Mainly caused by the inefficient operation of the utility, water losses cannot be considered exogenous parameters that affect the available technology of the firm through *g(Z,D)*.[39] Because network length is considered partially endogenous, we do not control for this variable either.

Appendix D: Final OLS Models for 2001

Table D-1

The significance of almost all the variables supports confidence in the variables included in the Final OLS Models for 2002. Regarding the value of the coefficients, the Population and Connections models are more stable than the Volumes model.

References

Berg, Sanford and Lin, Chen (2005) "Consistency in performance rankings: the Peruvian water sector", PURC Working Paper.

[39] Tupper and Resende (2004), however, apparently consider this variable as outside the control of the firm and construct measures of efficiency correcting for water losses.

Byrnes, Patricia; Grosskopf, Shawna and Hayes, Kathy (1986) "Efficiency and ownership: further evidence", Review of Economics and Statistics 68(2): 337-341.

Coelli, Tim; D.S. Prasada Rao, and Battese, George (1998) "An introduction to efficiency and productivity analysis", Boston: Kluwer Academic.

Corton, Maria Luisa (2003) "Benchmarking in the Latin American water sector: the case of Peru", Utilities Policy 11, 133-142.

Crain, W. Mark, and Zardkoohi, Asghar (1978) "A test of the property rights theory of the firm: water utilities in the United States", Journal of Law and Economics 21(2): 395-408.

Cubbin, John and Tzanidakis, George (1998) "Regression versus data envelopment analysis for efficiency measurement: an application to the England and Wales regulated water industry", Utilities Policy 7: 75-85.

Estache, Antonio and Kouassi, Eugene (2002) "Sector organization, governance and the inefficiency of African water utilities", World Bank Policy Research Working Paper N2890.

Estache, Antonio and Rossi, Martin (1999) "Comparing the performance of public and private water companies in Asia and Pacific region: what a stochastic costs frontier shows", World Bank Policy Research Working Paper.

Estache, Antonio and Rossi, Martin (2002) "How different is the efficiency of public and private water companies in Asia?", The World Bank Economic Review 16(1): 139-148.

Feigenbaum, Susan and Teeples, Ronald (1983) "Public vs. private water delivery: a hedonic cost approach", The Review of Economics and Statistics 65(4): 672-678.

Fox, William and Hofler, Richard (1986) "Using homothetic composed error frontiers to measure water utility efficiency", Southern Economic Journal 53(2): 461-477.

Kim, H. Youn (1987) "Economies of scale in multiproduct firms: an empirical analysis", Economica (new series) 54(214): 185-206.

Kumbhakar, Subal and Knox, Lovell (2000) "Stochastic frontier analysis", New York: Cambridge University Press.

Saal, David and Parker, David (2000) "The impact of Privatization and Regulation on the Water and Sewerage Industry in England and Wales: A Translog cost function approach", Managerial and Decision Economics 21(6): 253-268.

Saal, David and Parker, David (2001) "Productivity and price performance in the privatized water and sewerage companies of England and Wales", Journal of Regulatory Economics 20(1): 61-90.

Teeples, Ronald and Glyer, David (1987) "Cost of water delivery systems: specification and ownership effects," The Review of Economics and Statistics 69(3): 399-408.

Tupper, Henrique, and Resende, Marcelo (2004) "Efficiency and regulatory issues in the Brazilian water and sewage sector: an empirical study" Utilities Policy 12: 29-40.

Ministério das Cidades do Brazil (2002) "Visao Geral da Prestacao dos Servicos de Agua e Esgotos".

Table 1: Summary Statistics for the Basic OLS Models.

For the Volumes Model:

Variable	Obs	Mean	Std. Dev.	Min	Max
LN OPER COST	134	15.95228	2.006561	11.74275	21.38566
LN WAGE	134	9.741731	.6397651	6.343441	11.128
LN WATER VOL PROD	134	9.587815	1.823032	5.736572	14.76673
LN SEWER VOL COLL	134	8.296346	1.990508	3.072693	13.60176

For the Population Model:

Variable	Obs	Mean	Std. Dev.	Min	Max
LN OPER COST	152	15.81315	2.057066	11.33206	21.38566
LN WAGE	152	9.743298	.6214466	6.343441	11.128
LN POP SERV WATER	152	11.88342	1.761272	7.702105	16.8682
LN POP SERV SEWER	152	11.01836	1.959908	5.924256	16.63715

For the Connections Model:

Variable	Obs	Mean	Std. Dev.	Min	Max
LN OPER COST	152	15.81315	2.057066	11.33206	21.38566
LN WAGE	152	9.743298	.6214466	6.343441	11.128
LN WATER CONN	152	10.50077	1.708279	6.530878	15.48433
LN SEWER CONN	152	9.579431	1.93282	4.672829	15.18948

Table 2: Basic OLS Models for 2002.

Dependent variable: ln(Operating Cost)

Variable	Model		
	Volumes	Population	Connections
CONSTANT	0.51 (0.47)	-0.10 (-0.19)	0.68 (1.03)
WAGE	0.74*** (4.79)	0.34*** (4.62)	0.39*** (4.48)
WATER VOL PROD	0.65*** (5.51)		
SEWER VOL COLL	0.25*** (2.80)		
POP SERV WATER		0.95*** (25.32)	
POP SERV SEWER.		0.12*** (3.83)	
WATER CONN.			1.01*** (27.68)
SEWER CONN.			0.07** (2.57)
OBSERVATIONS	134	152	152
R^2	0.9234	0.9663	0.9668

Table 3: Summary Statistics for the Final OLS Models.

For the Volumes Model:

Variable	Obs	Mean	Std. Dev.	Min	Max
LN OPER COST	122	15.99308	2.039449	11.74275	21.38566
LN WAGE	122	9.771604	.5790253	7.179718	11.128
LN WATER VOL PROD	122	9.639481	1.851498	5.736572	14.76673
LN SEWER VOL COLL	122	8.309294	2.045749	3.072693	13.60176
PURCH WATER	122	.1393443	.3477335	0	1
LN WATER VOL MET	122	8.963044	1.871716	1.791759	14.05441
LN HHOLD WATER CONS	122	2.803977	.2925892	2.095135	4.249004

For the Population Model:

Variable	Obs	Mean	Std. Dev.	Min	Max
LN OPER COST	222	14.86881	2.299633	8.922459	21.38566
LN WAGE	222	9.588276	.6017808	7.004508	11.128
LN POP SERV WATER	222	11.17941	1.87881	7.166266	16.8682
SOUTHEAST	222	.3648649	.4824801	0	1
NORTHEAST	222	.2522523	.4352867	0	1
LN WATER VOL MET	222	7.856998	2.186113	-1.203973	14.05441

For the Connections Model:

Variable	Obs	Mean	Std. Dev.	Min	Max
LN OPER COST	222	14.86881	2.299633	8.922459	21.38566
LN WAGE	222	9.588276	.6017808	7.004508	11.128
LN WATER CONN	222	9.759834	1.877356	5.220356	15.48433
SOUTHEAST	222	.3648649	.4824801	0	1
LN WATER VOL MET	222	7.856998	2.186113	-1.203973	14.05441

Table 4: Final OLS Models for 2002.

Dependent variable: ln(Operating Cost)

Variable	Model		
	Volumes	Population	Connections
CONSTANT	5.14*** (3.98)	0.15 (0.21)	1.18 (1.59)
WAGE	0.37** (2.60)	0.33*** (3.52)	0.32*** (3.39)
WATER VOL PROD	0.48*** (2.99)		
SEWER VOL COLL	0.12*** (2.64)		
POP SERV WATER		0.90*** (22.33)	
WATER CONN.			0.97*** (27.78)
PURCH WATER	0.33** (2.21)		
HHOLD WATER CONS	-0.61*** (-3.99)		
SOUTHEAST		0.32*** (5.81)	0.25*** (4.93)
NORTHEAST		-0.13** (-2.12)	
WATER VOL MET	0.36** (2.05)	0.19*** (5.40)	0.13*** (4.41)
OBSERVATIONS	122	222	222
R^2	0.9582	0.9777	0.9789

Table 5: Final Models using Stochastic Cost Frontiers for 2002.

Dependent variable: ln(Operating Cost)

Variable	Model		
	Volumes	Population	Connections
CONSTANT	4.30*** (4.64)	0.14 (0.32)	0.90* (1.83)
WAGE	0.42*** (4.19)	0.33*** (5.94)	0.34*** (5.8)
WATER VOL PROD	0.54*** (8.02)		
SEWER VOL COLL	0.09** (2.34)		
POP SERV WATER		0.90*** (27.08)	
WATER CONN.			0.97*** (27.91)
PURCH WATER	0.24** (2.12)		
HHOLD WATER CONS	-0.64*** (-4.84)		
SOUTHEAST		0.32*** (5.66)	0.25*** (5.01)
NORTHEAST		-0.13** (-2.17)	
WATER VOL MET	0.33*** (4.03)	0.19*** (6.68)	0.13*** (4.43)
OBSERVATIONS	122	222	222
LOG LIKELIHOOD	-61.65	-77.35	-70.93

Table 6: Ranking of Regional firms for 2002 according to stochastic cost frontier models.

(from efficient to inefficient)

Volumes Model	Population Model	Connections Model
CAGECE/CE	**COPASA/MG**	**COPASA/MG**
CEDAE/RJ	**CAGECE/CE**	**CAGECE/CE**
COPASA/MG	**CESAN/ES**	SANEPAR/PR
SABESP/SP	**SABESP/SP**	**SABESP/SP**
CESAN/ES	SANEPAR/PR	**CESAN/ES**
SANEAGO/GO	COSANPA/PA	COMPESA/PE
COMPESA/PE	CEDAE/RJ	CASAN/SC
SIMAE/SC	SANESUL/MS	SANESUL/MS
SANEPAR/PR	COMPESA/PE	SANEAGO/GO
EMBASA/BA	SANEAGO/GO	CAEMA/MA
CAJ/RJ	CASAN/SC	CORSAN/RS
CASAN/SC	EMBASA/BA	SANEATINS/TO
CAGEPA/PB	SANEATINS/TO	EMBASA/BA
CORSAN/RS	CAEMA/MA	CAERN/RN
SANESUL/MS	CORSAN/RS	CAGEPA/PB
CAERN/RN	DESO/SE	SAAE/PR
DESO/SE	CAERN/RN	AGESPISA/PI
CAEMA/MA	CAGEPA/PB	COSANPA/PA
CAER/RR	CAJ/RJ	DESO/SE
COSANPA/PA	SAAE/ES	SAAE/ES
PROLAGOS/RJ	SAAE/PR	CEDAE/RJ
CASAL/AL	CAER/RR	CAJ/RJ
SANEATINS/TO	SIMAE/SC	SIMAE/SC
SAAE/ES	AGESPISA/PI	SIMAE/SC
CAESB/DF	SIMAE/SC	**CASAL/AL**
CAESA/AP	**CASAL/AL**	CAER/RR
	PROLAGOS/RJ	PROLAGOS/RJ
	CAESB/DF	**CAESB/DF**
	CAERD/RO	**CAESA/AP**
	CAESA/AP	CAERD/RO

Table A 1: Basic OLS Models for 2001

Dependent variable: ln(Operating Cost)

Variable	Model		
	Volumes	Population	Connections
CONSTANT	1.50 (1.15)	-0.37 (-0.53)	0.62 (0.78)
WAGE	0.60*** (3.36)	0.38*** (4.10)	0.40*** (3.94)
WATER VOL PROD	0.71*** (6.15)		
SEWER VOL COLL	0.21** (2.41)		
POP SERV WATER		0.91*** (21.96)	
POP SERV SEWER.		0.15*** (4.76)	
WATER CONN.			0.99*** (25.12)
SEWER CONN.			0.10*** (3.40)
OBSERVATIONS	114	132	133
R^2	0.9154	0.9659	0.9654

Table B 1: Mixed Basic OLS Models

Dependent variable: ln(Operating Cost)

Variable	Model		
	Vol/Pop	Vol/Conn	Pop/Conn
CONSTANT	0.15 (0.29)	0.90 (1.56)	0.38 (0.63)
WAGE	0.39*** (4.98)	0.41*** (5.28)	0.35*** (4.52)
WATER VOL PROD	0.04 (0.53)	0.01 (0.18)	
SEWER VOL COLL	0.10 (1.01)	0.12* (1.74)	
POP SERV WATER	0.88*** (10.48)		0.36* (1.67)
POP SERV SEWER.	0.02 (0.23)		0.14 (1.07)
WATER CONN.		0.96*** (11.73)	0.62*** (2.81)
SEWER CONN.		-0.04 (-0.56)	-0.04 (-0.33)
OBSERVATIONS	134	134	152
R^2	0.9688	0.9715	0.9689

Table B 2: Correlation coefficients between output variables.

	WATER VOL PROD	SEWER VOL COLL	POP SERV WATER	POP SERV SEWER.	WATER CONN.	SEWER CONN.
WATER VOL PROD	1					
SEWER VOL COLL	0.9625	1				
POP SERV WATER	0.9468	0.9310	1			
POP SERV SEWER.	0.9317	0.9692	0.9229	1		
WATER CONN.	0.9209	0.9130	0.9925	0.9289	1	
SEWER CONN.	0.8864	0.9362	0.8969	0.9915	0.9171	1

Table C 1: Basic OLS Models with all regional dummies

Dependent variable: ln(Operating Cost)

Variable	Model		
	Volumes	Population	Connections
CONSTANT	0.63 (0.61)	-0.42 (-0.73)	0.48 (0.67)
WAGE	0.74*** (4.96)	0.36*** (4.67)	0.41*** (4.56)
WATER VOL PROD	0.64*** (4.76)		
SEWER VOL COLL	0.26** (2.41)		
POP SERV WATER		1.04*** (28.66)	
POP SERV SEWER.		0.02 (0.79)	
WATER CONN.			1.07*** (28.54)
SEWER CONN.			0.02 (0.52)
NORTHEAST	-0.24 (-1.24)	-0.33*** (-2.67)	-0.23 (-1.58)
NORTH	-0.09 (-0.30)	-0.02 (-0.12)	0.19 (1.06)
SOUTHEAST	-0.15 (-0.77)	0.24** (1.99)	0.16 (1.08)
SOUTH	-0.12 (-0.63)	-0.01 (-0.11)	-0.08 (-0.55)

OBSERVATIONS	134	152	152
R^2	0.9243	0.9742	0.9715

Table C 2: Basic OLS Models with Northeast and Southeast dummies

Dependent variable: ln(Operating Cost)

Variable	Model		
	Volumes	Population	Connections
CONSTANT	0.51 (0.47)	-0.42 (-0.74)	0.43 (0.61)
WAGE	0.74*** (4.79)	0.36*** (4.65)	0.40*** (4.46)
WATER VOL PROD	0.65*** (5.51)		
SEWER VOL COLL	0.25*** (2.80)		
POP SERV WATER		1.04*** (29.05)	
POP SERV SEWER.		0.02 (0.80)	
WATER CONN.			1.08*** (28.44)
SEWER CONN.			0.003 (0.10)
NORTHEAST		**-0.32*** (-3.96)**	**-0.21** (-2.56)**
SOUTHEAST		**0.25*** (3.32)**	**0.19** (2.16)**
OBSERVATIONS	134	152	152
R^2	0.9234	0.9742	0.9709

Table C 3: Basic OLS Models with Purchased Water dummy

Dependent variable: ln(Operating Cost)

Variable	Model		
	Volumes	Population	Connections
CONSTANT	1.33 (1.49)	0.04 (0.07)	0.84 (1.26)
WAGE	0.66*** (5.32)	0.34*** (4.63)	0.38*** (4.47)
WATER VOL PROD	0.67*** (7.17)		
SEWER VOL COLL	0.20*** (3.01)		
POP SERV WATER		0.94*** (26.32)	
POP SERV SEWER.		0.11*** (3.59)	
WATER CONN.			1.01*** (27.99)
SEWER CONN.			0.07** (2.36)
PURCH. WATER	**0.72*** (3.40)**	**0.17 (1.46)**	**0.18 (1.56)**

OBSERVATIONS	132	150	150
R^2	0.9360	0.9671	0.9676

Table C 4: Basic OLS Models with Water Density

Dependent variable: ln(Operating Cost)

Variable	Model		
	Volumes	Population	Connections
CONSTANT	0.51 (0.47)	0.21 (0.25)	0.79 (0.99)
WAGE	0.74*** (4.79)	0.40*** (4.68)	0.46*** (5.42)
WATER VOL PROD	0.65*** (5.51)		
SEWER VOL COLL	0.25*** (2.80)		
POP SERV WATER		0.93*** (24.59)	
POP SERV SEWER.		0.14*** (4.27)	
WATER CONN.			0.99*** (27.60)
SEWER CONN.			0.09*** (3.23)
WATER CONN P/ KM		**-0.19*** **(-1.73)**	**-0.19*** **(-1.82)**
OBSERVATIONS	134	149	149
R^2	0.9234	0.9670	0.9682

Table C 5: Basic OLS Models with a Water Metering variable

Dependent variable: ln(Operating Cost)

Variable	Model		
	Volumes	Population	Connections
CONSTANT	3.01 (2.41)	0.76 (0.92)	0.96 (1.14)
WAGE	0.40*** (2.65)	0.31*** (2.97)	0.39*** (3.78)
WATER VOL PROD	0.38** (2.46)		
SEWER VOL COLL	0.09 (1.65)		
POP SERV WATER		0.80*** (11.21)	
POP SERV SEWER.		0.08** (2.50)	
WATER CONN.			0.89*** (12.41)
SEWER CONN.			0.05* (1.74)
WATER VOL MET.	**0.52*** (2.75)**	**0.20*** (2.61)**	**0.15** (1.99)**
OBSERVATIONS	131	146	146
R^2	0.9441	0.9673	0.9677

Table C 6: Basic OLS Models with Avg Water Household Consumption

Dependent variable: ln(Operating Cost)

Variable	Model		
	Volumes	Population	Connections
CONSTANT	3.27*** (3.79)	-0.43 (-0.57)	-0.11 (-0.15)
WAGE	0.68*** (4.80)	0.40*** (4.72)	0.47*** (5.39)
WATER VOL PROD	0.66*** (5.51)		
SEWER VOL COLL	0.24*** (2.63)		
POP SERV WATER		0.93*** (22.80)	
POP SERV SEWER.		0.13*** (3.86)	
WATER CONN.			1.00*** (25.49)
SEWER CONN.			0.08*** (2.60)
HHOLD WATER CONS	**-0.81*** (-6.19)**	**-0.04 (-0.32)**	**0.05 (0.44)**

OBSERVATIONS	126	144	144
R^2	0.9451	0.9658	0.9697

Table C 7: Final OLS Models for 2002

Dependent variable: ln(Operating Cost)

Variable	Model		
	Volumes	Population	Connections
CONSTANT	5.14*** (3.98)	0.15 (0.21)	1.18 (1.59)
WAGE	0.37** (2.60)	0.33*** (3.52)	0.32*** (3.39)
WATER VOL PROD	0.48*** (2.99)		
SEWER VOL COLL	0.12*** (2.64)		
POP SERV WATER		0.90*** (22.33)	
WATER CONN.			0.97*** (27.78)
PURCH WATER	0.33** (2.21)		
HHOLD WATER CONS	-0.61*** (-3.99)		
SOUTHEAST		0.32*** (5.81)	0.25*** (4.93)
NORTHEAST		-0.13** (-2.12)	
WATER VOL MET	0.36** (2.05)	0.19*** (5.40)	0.13*** (4.41)
OBSERVATIONS	122	222	222
R^2	0.9582	0.9777	0.9789

Table D 1: Final OLS Models for 2001

Dependent variable: ln(Operating Cost)

Variable	Model		
	Volumes	Population	Connections
CONSTANT	3.11** (2.39)	-0.12 (-0.23)	0.88* (1.73)
WAGE	0.50*** (3.34)	0.36*** (5.76)	0.35*** (5.64)
WATER VOL PROD	0.19 (1.51)		
SEWER VOL COLL	0.06*** (2.74)		
POP SERV WATER		0.91*** (18.60)	
WATER CONN.			0.99*** (20.46)
PURCH WATER	0.21** (2.26)		
HHOLD WATER CONS	-0.40*** (-2.67)		
SOUTHEAST		0.29*** (4.96)	0.28*** (5.16)
NORTHEAST		-0.21*** (-3.31)	
WATER VOL MET	0.74*** (5.22)	0.17*** (3.44)	0.10** (2.00)

OBSERVATIONS	94	186	190
R^2	0.9681	0.9785	0.9790