# DEA AND ITS USE IN THE REGULATION OF WATER COMPANIES[1]

by

EMMANUEL THANASSOULIS

Aston Business School, University of Aston, Birmingham B4 7ET, UK

E-MAIL E.THANASSOULIS@ASTON.AC.UK.

***Presented at EURO XVI, Brussels, Belgium, July 1998***

**ABSTRACT**

The paper begins with an introduction to the basic principles of Data Envelopment Analysis (DEA). DEA is a linear programming - based method for assessing the productive efficiencies of operating units such as bank branches, sales outlets, schools or individuals. The paper then goes on to describe the use of DEA in the regulatory framework. Regulation, employed to safeguard the public interest, is increasingly playing an major role in Great Britain and other countries in the aftermath of the privatisation of publicly owned companies including utilities which still enjoy a good degree of monopoly power. The paper gives an account of the use of DEA to estimate potential cost savings at water companies in the context of the price review conducted by the regulator of water companies in England and Wales in 1994. It also highlights certain generic issues arising in the use of DEA in the regulatory context. The paper should prove of interest both to those who want to know about DEA as a tool in general and to those interested in efficiency measurement under regulation.

*Keywords: Data Envelopment Analysis, Performance Measurement, Regulation, Water Industry*

[1] The author is grateful to the regulator of UK water companies (OFWAT) for permission to publish extracts from analyses undertaken on its behalf. The views expressed in this document are those of the author and no representation is being made that they are necessarily shared by OFWAT. Any shortcomings in the paper are entirely the responsibility of the author.

# 1. INTRODUCTION

DEA, developed by Charnes et al. (1978), is a general purpose, linear programming - based method for assessing the productive efficiencies of operating units such as bank branches, schools and hospitals which perform a given function. Each unit transforms a set of ‘resource’ factors such as labour and capital (referred to as ‘inputs’) into a set of desirable outcomes such as goods or services, referred to as ‘outputs’. In this paper we outline the use of DEA in a regulatory framework.

Regulation has taken on an enhanced role in the UK economy in the wake of the massive programme of privatisations of publicly owned assets in the 1980’s. Privatisations are generally intended to reinforce market competition but such competition is not perfect, at least not so far as many utilities are concerned. For example most houses and businesses can only be supplied with water through their connection to the distribution mains even if there is supplier competition so far as the water itself is concerned. The company controlling the distribution mains has monopolistic powers in this situation. The regulatory systems put in place by the UK government aim to counter such monopolistic powers. In the case of UK water companies the regulator is the Office of Water Services, known as OFWAT.

In 1994 OFWAT conducted the first Periodic Review of the water companies in England and Wales with a view to setting price caps for a ten year period post 1993/4. As part of this review the question was addressed as to what operating cost savings are in principle feasible at water companies. DEA was one of the methods used to address this question and this paper outlines the use of DEA in this context.

The paper is divided into two parts. Part I gives an introduction to DEA. Part II outlines the use of DEA by OFWAT in the framework of the Periodic Review of water companies in 1994 and draws out some generic issues pertaining to the use of DEA in regulation.

# 2. PART I: INTRODUCTION TO DATA ENVELOPMENT ANALYSIS

## 2.1. MEASURES OF EFFICIENCY UNDER DEA

DEA is applied to *units of assessment* such as the branches of a bank or schools. The unit of assessment is normally referred to as a *decision making unit* (DMU). A DMU converts "inputs" into "outputs" in a process depicted in Figure 1.

**Figure 1 about here please**

The identification of what are the inputs and the outputs in an assessment of DMUs is as difficult as it is crucial. For example if a Police Force is to be a unit of assessment then a possible set of input - output variables (see Thanassoulis 1995) is as follows.

**Table 1: Input - output variables for Police Forces**

| **Inputs** | **Outputs** |
|---|---|
| - violent crimes | - violent crime clear ups |
| - burglaries | - burglary clear ups |
| - Other crimes | - Other crime clear ups |
| - Officers | |

The assessment reflects the 'efficiency' of the Forces in securing crime clear ups given the level of crime they each have to deal with and the manpower they have been given by governing authorities.

In measuring the efficiency of a DMU we need to know whether the unit could have secured more output for its input levels or could have used less input for its output levels. We define a *Pareto - efficient* or **DEA - efficient** DMU in a case where the DMU uses $m \geq 1$ inputs to secure $s \geq 1$ outputs in either one of the following orientations:

*Output orientation:* A DMU is *Pareto - efficient* if it is not possible to raise anyone of its output levels without lowering at least another one of its output levels and/or without increasing at least one of its input levels.

*Input orientation:* A DMU is *Pareto - efficient* if it is not possible to lower anyone of its input levels without increasing at least another one of its input levels and/or without lowering at least one of its output levels.

Two measures of efficiency, relating respectively to the output and input orientation above, are as follows:

| | |
|---|---|
| *Technical output efficiency* | The technical output efficiency of a DMU is the inverse of the maximum factor by which its output levels could be expanded while its input levels do not rise. |
| *Technical input efficiency* | The technical input efficiency of a DMU is the maximum factor by which its input levels could be contracted while its output levels do not fall. |

Both of these measures relate to equiproportionate contraction of input or expansion of output levels and for this reason they are known as *radial* measures of efficiency. The use of the prefix 'technical' in the above efficiency measures is because they relate to technical transformation of inputs into outputs without regard to input or output prices. *Allocative* or *Price* efficiency measures which do reflect input prices can be found in Farrell (1957).

Figure 2 illustrates the foregoing measures of efficiency:

**Figure 2 about here please**

Figure 2 depicts the case where DMUs produce a single output using a single input. The curve OD is the locus of maximum output levels attainable for given input levels. OD is thus the *efficient boundary* of the *production space* located between the input axis and OD. What we mean by the term 'production space' here is the set of feasible input - output combinations, which is generally referred to in DEA as the *Production Possibility Set* (PPS). The points below the curve OD either use more input for given output or produce less output for given input compared to points on OD. We assume excess input or output is 'freely disposable' and so if the points on OD are feasible so will the points below OD, this being known as the assumption of *free disposal*.

Let us now consider DMU A in relation to the definition of Pareto - efficiency and the measures of efficiency introduced earlier. Clearly it is possible to expand its output level to D without raising its input level and so DMU A is not Pareto - efficient. The technical output efficiency of

DMU A was defined above as $\frac{1}{\frac{OB}{OH}} = \frac{OH}{OB}$. Its technical input efficiency was defined as $\frac{OF}{OG}$.

These two measures of efficiency are not in general equal except under *constant returns to scale* (CRS). (CRS and *variable returns to scale* (VRS) are discussed later.) Under CRS the efficient boundary OD in Figure 2 would have been a straight line through the origin.

## 2.2. MEASURING INPUT EFFICIENCY USING DATA ENVELOPMENT ANALYSIS

An essential step in measuring technical efficiency in Figure 2 was the construction of the production possibility set, (PPS). How this is done in DEA is illustrated using a graphical example.

**Example**

A bank has four branches. The table below shows the number of hours of staff time used weekly per 1000 accounts in existence at the branch. What is the efficiency of branch 1?

**Table 2: Input – output variables of bank branches**

| **BRANCH** | **Supervisory hours** (*Input 1*) | **Trainee hours** (*Input 2*) | **Accounts (000)** *(Output)* |
|---|---|---|---|
| 1 | 2 | 3 | 1 |
| 2 | 4 | 1 | 1 |
| 3 | 2 | 2 | 1 |
| 4 | 1 | 4 | 1 |

We construct the PPS in DEA from the observed input - output correspondences or DMUs using the following basic assumptions (see Banker et al. (1984, p.1081) for a formal statement of these assumptions):

- Interpolation between observed input - output correspondences leads to observable input - output correspondences;
- Inefficient transformation of inputs to outputs is possible;
- The efficient transformation of inputs to outputs is characterised by constant returns to scale;

- The PPS is the smallest set meeting the foregoing assumptions and containing all input - output correspondences observed at DMUs.

  **Figure 3 about here please**

The construction of a set of feasible points using these assumptions is shown in Figure 3. Point L is obtained as the interpolation of branches 3 and 4 so that L = 0.5 Branch 3 + 0.5 Branch 4. The interpolation assumption means that 1.5 supervisory and 3 trainee hours, <u>though observed at no branch</u>, are capable of handling 1000 accounts. Moreover, using the inefficiency assumption we deduce that all input - output levels which satisfy

*Supervisory hours ≥1.5*

*Trainee hours ≥3 and*

*Accounts ≤ 1*

are also feasible. The shaded area in Figure 3 shows these feasible points for the case Accounts = 1. Figure 4 generalises this process to construct the PPS based on the four branches. B2B3B4 and the space above and to its right in Figure 4 contains all input levels capable of handling 1000 accounts which we can construct from the four branches. This is the 'Production Possibility Set', <u>PPS</u>.

**Figure 4 about here please**

The radial efficiency measure defined earlier if applied to Branch 1 requires the contraction of its input levels while retaining their mix of 2 supervisory to 3 trainee hours. The mix of Branch 1 defines the line OB1 in Figure 5. Point M in Figure 5 shows the lowest input levels Branch 1 could have used to handle 1000 accounts while keeping to its input mix. The input levels at M are 2.572 for trainee and 1.714 for supervisory hours. Thus using the definition of technical input efficiency given earlier we deduce that the relative efficiency of Branch B1 is $\frac{1.714}{2} = \frac{2.572}{3}$ = 0.857.

**Figure 5 about here please**

All points along the boundary of the PPS cannot have trainee and supervisory hours reduced <u>simultaneously</u> and so their radial efficiency is 1. However, points on a vertical extension from

B4 or horizontal extension from B2 can have one input, but not both, reduced.  Thus points on B2B3B4 are Pareto or DEA - efficient but not so for any other boundary points.

It is evident from Figure 5 that the DEA efficiency measure of 0.857 obtained for B1 is specifically with reference to Branches 3 and 4.  Branches B3 and B4 are known as the **efficient referents** or **efficient peers** of Branch 1.  The input-output levels of the efficient comparator 'branch' M in Figure 5 represent **efficient targets** for Branch 1.

## 2.3. MEASURING DEA EFFICIENCY: A LINEAR PROGRAMMING APPROACH

The following linear programming model simulates the graphical estimation of the DEA - efficiency of Branch 1 carried out in the previous section.

$$
\begin{aligned}
&\text{Min } Z_1 \\
&\text{Such that} \\
&2\,Z_1 = 2\,\lambda 1 + 4\,\lambda 2 + 2\,\lambda 3 + \lambda 4 \\
&3\,Z_1 = 3\,\lambda 1 + \lambda 2 + 2\,\lambda 3 + 4\,\lambda 4 \qquad\qquad [M1] \\
&1 = \lambda 1 + \lambda 2 + \lambda 3 + \lambda 4 \\
&\lambda 1, \lambda 2, \lambda 3\ \lambda 4 \geq 0\ .
\end{aligned}
$$

Any set of feasible $\lambda$ values yields a point on MB1 in Figure 5.  The minimum value of $Z_1$ in [M1] corresponds to point M in Figure 5, reflecting the lowest proportion to which the input levels

(2, 3) of Branch 1 can be lowered while output is maintained at 1000 accounts.  The minimum value of $Z_1$ in [M1]  is $Z_1{}^* = 0.857$, the same as the efficiency rating obtained graphically in Figure 5.

The model in [M1] is a special case of the general purpose DEA model developed by Charnes et al. (1978) which is as follows.  Let us have $N$ DMUs $(j = 1, \ldots, N)$ using $m$ inputs to secure $s$

outputs. Let us denote $x_{ij}$ , $y_{rj}$ the observed level of the ith input and rth output respectively, at DMU j. The following linear programming model is solved to ascertain whether DMU $j_0$ is 'DEA - efficient' and measure its efficiency:

$$h_{j_0} = \text{Min } k_0 - \varepsilon \left[\sum_{i=1}^{m} S_i^- + \sum_{r=1}^{s} S_r^+\right]$$

$$\text{S. t. } \sum_{j=1}^{N} \lambda_j x_{ij} = k_0 x_{ij_0} - S_i^- \qquad \text{i= 1...m}$$

$$\sum_{j=1}^{N} \lambda_j y_{rj} = S_r^+ + y_{rj_0} \qquad \text{r = 1...s} \qquad \text{[M2]}$$

$$\lambda_j \geq 0, j = 1 \ldots N, S_i^-, S_r^+ \geq 0 \ \forall\ i \text{ and } r,$$

$k_0$ free. $0<\varepsilon<<1$ is a non - Archimedean infinitesimal.

Model [M2] works as follows. For a given set of feasible λ values the LHSs of the input and output related constraints specify a production point within the PPS. The model seeks a PPS point which offers at least the output levels of DMU $j_0$ while using as low a proportion of its input levels as possible. Let the superscript * denote optimal values.

- **DMU $j_0$ is <u>DEA - efficient</u> if and only if $k_0^*$ = 1 and $S_r^{+*} = 0, r = 1 \ldots s,\ S_i^{-*} = 0,\ i = 1 \ldots m.$**
- **$k_0^*$ is a measure of the radial <u>DEA efficiency</u> of DMU $j_0$ .**

Model [M2] assesses efficiency in a production context. Its dual assesses efficiency in a value context. The dual to [M2] is as follows:

$$h_{j_0} = \text{Max} \sum_{r=1}^{s} u_r y_{rj_0}$$

$$\textit{subject to} \sum_{i=1}^{m} v_i x_{ij_0} = 1$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0 \qquad \text{j = 1... } j_0\text{...N} \qquad \text{[M3]}$$

$$u_r \geq \varepsilon, \text{ r = 1 ... s, } v_i \geq \varepsilon, \text{ i = 1 ... m.}$$

The variables $u_r$ and $v_i$ are dual to the constraints relating to the rth output and ith input in [M2] respectively. By virtue of duality models [M2] and [M3] yield the same efficiency rating $h_{j_0}$ in respect of DMU $j_0$. However, the efficiency rating in [M3] can be seen in a value rather than production context. The optimal value $u_r^*$ of $u_r$ can be seen as **the imputed value** per unit of output r. Similarly, the optimal value $v_i^*$ of $v_i$ can be seen as the imputed value per unit of input i. When we view $u_r^*$ and $v_i^*$ as imputed values in this way the efficiency measure $h_{j_0}$ of DMU $j_0$ yielded by [M3] is **the ratio of the total imputed value of its output levels to that of its input levels**. (See Charnes et al. (1978) or Thanassoulis (1997a) for the derivation of this ratio.) The use of ratios to reflect efficiency has been commonplace since well before DEA. Indeed, practitioners have often used prior weights to reduce multiple outputs to a single 'weighted output' and multiple inputs to a single 'weighted input' to arrive at a measure of performance. What makes the DEA measure of efficiency in model [M3] different are two key facts:

- No prior values of inputs and outputs are imposed and;
- The input - output values are DMU - specific, chosen to maximise the efficiency rating of the respective DMU.

For more details on equivalences between the value and the production contexts of DEA assessments of performance and their practical implications see Thanassoulis (1997a).

## 2. 4. ALLOWING FOR VARIABLE RETURNS TO SCALE

The models developed so far are based on the assumption that efficient input - output levels are characterised by constant returns to scale (CRS). **Under local CRS, if DMU A is efficient and its input levels are scaled by (1+ a) where |a| <<1, then its output levels will also be scaled by (1+ a).** The CRS assumption made so far in measuring DEA efficiency was relaxed by Banker et al. (1984). The different efficiencies resulting under constant and variable returns to scale for a DMU are illustrated in Figure 6. The graph depicts the input output levels of DMUs

A-D. The DMUs use a single input to produce a single output. Under CRS the efficient boundary is OA and its extension. Under CRS the efficient projection of DMU B would be at F. However, DMU B has lower output level than A and under VRS we cannot extrapolate the output to input ratio from DMU A to DMU B. We therefore construct a different production space permitting interpolation but not extrapolation of output - input combinations observed. **The VRS boundary in Figure 6 is CAD**. For example DMU B is not efficient under variable or constant returns to scale. Its input efficiency under VRS is $\frac{GE}{GB}$. Its input efficiency under CRS is $\frac{GF}{GB}$. Note that the VRS boundary is contained within the CRS and so **efficiencies estimated under VRS are always at least as high as those estimated under CRS.**

- Efficiency under VRS is termed **pure technical efficiency**.
- Efficiency under CRS is termed **overall technical efficiency**.

**Figure 6 about here please**

In the general case the VRS <u>pure technical input efficiency</u> of DMU $j_0$ is yielded by the optimal objective function value of model [M2] after we add the 'convexity' constraint $\sum_{j=1}^{N} \lambda_j = 1$**.** The pure technical input efficiency of a DMU does <u>not</u> necessarily equal its pure technical output efficiency. The ratio $\frac{CRS\ Efficiency}{VRS\ Input\ Efficiency}$ is the <u>**input scale efficiency**</u> of the DMU**.** This is a measure of the part of the inefficiency of a DMU which can be attributed to its operating away from the "most productive scale size" ( Banker (1984). The DEA model solved can be used to identify whether a DMU on the VRS efficient boundary operates under constant, increasing or decreasing returns to scale hold, (Banker and Thrall (1992)).

# 3. PART II: AN OUTLINE OF THE USE OF DEA IN THE REGULATION OF UK WATER COMPANIES

## 3. 1. THE REGULATION FRAMEWORK OF THE UK WATER INDUSTRY

The regulation mechanism used by OFWAT, the regulator of UK water companies, is the **Price Cap**. It takes the form

$$PC_t^j = \mathrm{RPI}_t + K_t^j \qquad (1),$$

where $PC_t^j$ is change to the average annual charges company j is permitted to make in year t, ‘$\mathrm{RPI}_t$’ is the change in the Retail Prices Index from year t-1 to year t and $K_t^j$ is a **company - specific factor**, determined by the regulator for year t. The regulator’s sole concern in this process is the estimation of the company and year specific K- factors in (1) which give effect to the price - caps imposed on the companies. The company specific factors are estimated to reflect the balance between price increases the company is permitted due to mandatory improvements to water quality and the environment and price reductions which the company must effect through operating efficiency gains. DEA was one of the tools, but not the only one, employed by OFWAT to estimate such efficiency gains. We draw here from the DEA - based estimations of savings in the water distribution and sewerage functions to illustrate the use of DEA. See also Thanassoulis (forthcoming, 1997b).

## 3.2. THE CREATION OF UNITS OF ASSESSMENT BY SUBDIVIDING WATER COMPANIES

Water companies are too complex to constitute a unit of assessment. Moreover, they differ substantially in the services they deliver. At the time of the first Periodic Review by OFWAT (1994) some ten companies delivered clean water and collected sewerage. Such companies are referred to as Water and Sewerage Companies or WASCs. A further 22 companies had only clean water functions and they are referred to as Water only Companies of WoCs. OFWAT decided on units of assessment at company function level. Figure 7 shows the functions delineated for this purpose. Assessment by functions makes it possible to use fewer input - output variables than would be required for units of assessment at company level. The latter would seriously hamper discrimination on efficiency given the small number of companies under assessment.  The potential savings estimated at function level were then aggregated to estimate potential savings at company level.

**Figure 7 about here please**

The 10 WASCs accounted for some 75% of water delivered in England and Wales.  **OFWAT decided that only WASCs will be permitted as referent or efficient peer companies.**  WoCs account for a minority of water delivered and their relatively low level of assets and simpler organisational and operating structures made them intuitively unsuitable role models for the much larger WASCs.  This decision means that a conservative view is being taken on potential savings at both WASCs and WoCs.  Figure 8 illustrates the point. It shows the observed LENGTH and PROPERTIES per unit of OPEX at each company.  The outer boundary enveloping all companies corresponds to the case when WoCs are permitted to be referent companies.  Though the boundary need not have contained WoCs exclusively it nevertheless turned out to be the case.  The inner boundary consisting of the solid thin and thick lines corresponds to the case when only WASCs are permitted to define the DEA efficient boundary.

**Figure 8 about here please**

The DEA - efficiency rating $h_{j_0}$ which model [M2] would yield in respect of company A is $\frac{OA}{OB}$ when WoCs are not permitted to define the efficient boundary.  In contrast, when WoCs are permitted to define the efficient boundary company A has DEA - efficiency rating $h_{j_0} = \frac{OA}{OC}$.

Clearly the DEA - efficiencies will always be higher with reference to the inner rather than the outer boundary.  This in turn means that the estimated potential savings of all companies will be lower when the referent boundary is the inner one, corresponding to the case when only WASCs are permitted to define the efficient boundary.

### 3.3. CHOOSING THE INPUT - OUTPUT VARIABLES

This is a most important initial stage in any DEA assessment.  The function of water delivered is used here to illustrate the approach followed.  (For more details see Thanassoulis (forthcoming).)  It covers the conveying of water taken from the water treatment works to the clients.  The inputs should reflect the resources used and the outputs the volume and quality of activities encapsulated in the function being modelled.  This function in its essence concerns the pumping of water to high storage points and delivering mostly by force of gravity. The key assets used are pumps, manpower and the distribution main. OFWAT had prior to this analysis, and in the context of the econometric analyses of company efficiencies, identified factors which 'best' explain water distribution OPEX. (See OFWAT 1993, 1994a.) Drawing on these prior analyses an initial list of potential input - output factors for the DEA assessment was compiled, reproduced in Table 3.

**Table 3: Potential input - output variables in the distribution of water.**

| **Input** | **Potential Outputs** |
|---|---|
| OPEX | PROPERTIES |
| | LENGTH OF MAIN |
| | WDELA |
| | MEASN |
| | REMWDA and |
| | BURSTS |

OPEX stands for OPerating EXpenditure and includes all variable resource expended in conveying the water from the water treatment works to the customers, except for power costs. Power costs relate almost exclusively to pumping the water. OFWAT assessed companies separately on their cost efficiency in pumping, (see OFWAT 1994a Appendix 3).

PROPERTIES reflects the number of connections served by a company and the LENGTH OF MAIN reflects their geographical dispersion. Water delivered WDELA to clients consists of a metered and a non - metered component. Household supplies were not generally being metered at the time of the assessment while those to business generally were. The two variables of MEASN and REMWDA break down water delivered respectively into that which is measured and the remainder which is estimated. Finally, BURSTS is a potential output because it reflects expenditure incurred in repairs to mains bursts. The outputs collectively explained in large measure OPEX variations across companies. The potential output variables in Table 3 are highly correlated as Table 4 shows.

**Table 4: Correlation coefficients**

| | WDELA | MEASN | PROPS | LENGTH | BURSTS |
|---|---|---|---|---|---|
| **WDELA** | | | | | |
| **MEASN** | **0.981** | | | | |
| **PROPS** | **0.995** | **0.984** | | | |
| **LENGTH** | 0.941 | 0.972 | 0.951 | | |
| **BURSTS** | 0.812 | 0.876 | 0.835 | 0.897 | |
| **REMWDA** | **0.997** | 0.963 | **0.989** | 0.917 | 0.777 |

The largest of the correlation coefficients are highlighted. It was decided to construct subsets of the outputs which might be used instead of the full set of outputs and observe the nature of any differences in the assessments of companies the subsets yield. Table 5 shows the initial three subsets of outputs constructed.

**Table 5: Three potential output sets for assessing DEA efficiencies in water distribution**

| Set | Outputs |
|---|---|
| 1: | PROPERTIES, LENGTH OF MAIN and WDELA |
| 2: | PROPERTIES and LENGTH OF MAIN |
| 3: | LENGTH OF MAIN and WDELA |

Model [M2] was solved in respect of each one of the 32 companies, and in respect of each one of the three output sets in Table 5. WoCs were not permitted as referent units throughout. Some 25 companies changed rank across the three output sets by no more than 2 places. A further four changed rank by only 3 to 5 places.

Only three companies changed rank on relative efficiency substantially across the output sets in Table 5. Two of them changed rank by 16 and 9 places respectively and were relatively more efficient when WDELA replaced PROPERTIES as an output. By implication they deliver unusually large amounts of water given the number of properties they serve. On closer inspection it was found that this was because they delivered an unusually large component of measured water. Indexing at 100 the largest proportion of measured water delivered the median index value across the industry was 55.5 but the two companies above had index values of 100 and 88.60 respectively, the two largest index values in the industry.

The opposite was the case with a third company whose rank on efficiency worsened by 11 places when water delivered was used as an output variable instead of properties served. The company had by far the lowest proportion of measured water delivered. Measured water is delivered to businesses in large volumes per client, and therefore should reflect lower expenditure than the same volume of water delivered to households. Thus the output set {LENGTH, PROPERTIES } should give the more accurate reflection of company cost efficiency among the output sets in Table 5.

Next, the impact of splitting WDELA into water delivered measured to non - households (MEASN) and the remainder of water delivered, (REMWDA) and of including BURSTS was assessed. The output sets used were those in Table 6.

**Table 6: Further potential output sets for assessing water distribution**

| **Set** | **Outputs** |
|---|---|
| 1: | PROPERTIES, LENGTH OF MAIN |
| 2: | PROPERTIES, LENGTH OF MAIN and BURSTS |
| 3: | PROPERTIES, LENGTH OF MAIN, MEASN and REMWDA |

As might have been expected, companies identified earlier as respectively benefiting or suffering on efficiency rating from exceptionally large or low proportions of measured water delivered changed efficiency rank substantially in going from output set 1 or 2 to output set 3 in Table 6. Otherwise company ranking on efficiency remained stable under the three output sets. Only one company changed efficiency rank substantially between output set 2 and either of the other two sets of outputs in Table 6. The company had by far the largest number of bursts per Km of main. Its data was deemed atypical and bursts were excluded as a potential output variable.

**In light of the foregoing observations the decision was made to adopt {PROPERTIES, LENGTH, WDELA} as the output set to be used for the analysis.** This gives similar results to the output set {PROPERTIES, LENGTH} which gives probably the fairer reflection of company efficiencies out of the output sets contained in Tables 5 and 6. However, the variable WDELA was added because it does give companies which deliver large volumes of water for the number of properties they serve ‘the benefit of the doubt’. That is it requires lower savings of them than would be the case if water delivered had not been used as an output variable.

### 3.4. USE MADE OF THE DEA ASSESSMENT RESULTS BY OFWAT IN 1994

Once the foregoing set of three output variables was decided upon the efficiency ratings and potential savings in OPEX on water distribution were computed using Model [M2]. The potential OPEX savings SAV $j_0$ at company $j_0$ are

SAV $j_0$ $= (1- k_0^*)$ OPEX $j_0$ (2),

where OPEX $j_0$ is the OPEX of company $j_0$ and $k_0^*$ is the optimal value of $k_0$ in [M2].

Assessments similar to that outlined above for water distribution were also carried out for water resources and treatment, the stage the water goes through before being distributed. (DEA assessments were also carried out to estimate potential savings in the sewerage function Thanassoulis (1997b).) The precise use of the results from the two approaches is confidential to OFWAT. However, from published accounts, (see (OFWAT (1995), p. 414)) it is found that the DEA estimates of efficient cost levels in distribution and resources and treatment

> *"were added to overall average* (clean) *water business activities costs and the result divided by the actual distribution, treatment and business costs to give an overall* (DEA - based) *efficiency ratio* (of clean water operations). *In most cases the results* (on company efficiency on clean water operations) *were similar to those of the regressions. If they were significantly better, the Director* (of OFWAT) *moved the company up one band* (on efficiency in clean water).*"* (Bracketed text has been added by way of explanation of the background to the quotation given in italics.)

Thus the DEA results were used by OFWAT in conjunction with independent OLS regression - based results to arrive at a ranking of companies on efficiency. Once the ranks on efficiency were obtained, OFWAT took further factors into account such as the quality of customer service provided by each company and its strategic plans before arriving at the final price determinations.

While a direct link between the DEA applications and the final OFWAT price determinations of 1994 is difficult to establish, their potential impact is very high in monetary terms. The water industry cost the public in England and Wales in 1993/4, at current prices about 4,000 million US dollars. The price determinations were to last for up to 10 years, impacting costs of the order of 40,000 US dollars. Thus even minor differences attributable to the DEA analyses in the final price determinations can have a very substantial financial impact both for the public and the companies. (In the event OFWAT announced a new round of price determinations for 1999).

## 3.5. GENERAL ISSUES IN THE USE DEA IN THE REGULATORY CONTEXT

Certain issues which are not fully resolved are especially relevant to the use of DEA in the regulatory framework.

*Returns to scale assumptions*

It can be argued that it is in the public interest that assessments in the regulatory context should always be under constant returns to scale so long as scale size is controllable by management, irrespective of the true nature of returns to scale at which their firm happens to operate. This would ensure that a firm operating through choice at an uneconomic scale size will not be permitted to pass on to the public cost inefficiencies. Figure 9 illustrates the point.

Under CRS firm P would be reflected at P1 while under VRS at P2. P1P2 represents the cost incurred by the firm for not operating at the most productive scale size, represented by firm E. Firm P should not be permitted to pass this cost on to its clients if it can control scale size.

**Figure 9 about here please**

There are, however, counter - arguments to the foregoing statement. Firstly, scale size may be dependent on such contextual variables as population served, dispersion of population and so on, and so not controllable by companies. The regulator may though control scale size indirectly in such cases by permitting mergers and acquisitions of regulated firms. Secondly, where regulated entities are the outcome of privatisations of publicly owned assets as is the case with UK utilities, the scale of assets (e.g. water treatment works, water mains ) each company inherited on privatisation is clearly beyond managerial control. Where assets have long lives (as in the case of water companies) in the short to medium term management cannot change the scale of operation to exploit returns to scale without incurring unjustifiable capital costs. There is an issue, however, as to whether the regulator should incorporate in the price determinations an element which encourages firms to move to most productive scale size.

*Technologies operated*

Regulators often assess cost efficiency taking as given the technology being operated by the regulated firms. For example in 1994 OFWAT sought efficiency savings taking as given the treatment processes for clean water and sewage used by each water company. This approach obviously cannot identify any cost inefficiency attributable to using cost - ineffective technologies. For example a certain quality of clean water may be attainable by a number of different processes, some more cost effective than others. To the extent that the technologies operated are under managerial control the regulator should not permit costs incurred through using uneconomic technologies to be passed on to the public. This argument is similar to that for using constant returns to scale to assess companies where scale size is under managerial control.

However, as in the case of scale size, technologies operated are in large measure inherited and not under managerial control in the short to medium term. The question remains, however, as to whether the regulator should factor into price determinations incentives for companies to adopt more cost - efficient technologies. (See also Thanassoulis 1997b on this point.)

*Factoring in some inefficiency*

If the regulator uses a boundary method such as DEA to estimate the potential cost savings a company would have to outperform the boundary, that is the most cost-efficient companies in order to make profits beyond those made by the boundary companies. This can be very demanding especially for less efficient companies and could prove detrimental to their longer term financial viability. OFWAT appears to have been mindful of this and it reports that in the 1994 Periodic Review it set the price limits to

> "*Bring most companies about half way from their existing cost levels to the costs of the more efficient companies. This takes account of the uncertainties involved in identifying an efficiency frontier, and is also designed to leave all companies incentives to achieve additional savings over and above those reflected in price limits*". (OFWAT 1994b, p. 31).

*Assessing complex operating units by parts*

It is frequently the case in the regulatory context that the number of companies involved is very small. This is the case with water and especially with sewerage companies in the UK of which there are only 10. Comparative performance measurement methods, such as DEA, work better the larger the number of comparative units. Regulated companies are often very complex entities and their aggregate set of operations would require many variables if it is to be reflected accurately. However, the larger the number of variables needed to reflect the operations of the companies the larger the number of companies needed to discriminate on their relative efficiency.

One way to increase the number of comparator units is through the use of panel data, treating each unit as a distinct comparative entity in each unit of time. The approach does depend on a relatively stable technology to make comparisons of performance across time meaningful. A second approach, and that used by OFWAT in 1994, is to divide the complex entities which are few in number (e.g. water companies) into self contained homogeneous parts and make the parts created in this manner the units of assessment.

For example OFWAT used a *two - stage decomposition* of water companies in order to increase the number of units of assessment. The two - stage decomposition was as follows:

Stage I: Decomposition of each company into major functions;

Stage II: Decomposition of each function into self contained operating units.

In the Stage I decomposition each company was decomposed into seven functions as can be seen in Figure 7. The Stage II decomposition was applied only to some functions and it entailed creating several assessment units relating to a given function of a company. For example in the case of sewerage (see Thanassoulis 1997b and Figure 7) some 60 *Areas* were created as the units of assessment, each company having a number of sewerage Areas. Similarly, Sewage Treatment was assessed at works level, different assessment models being adopted for "Large" and for "Small" works. (See OFWAT 1994a).

Where the decomposition of large complex entities into self contained units is feasible it is a reasonable response to the problem of too few and very complex units to assess. There are, however, certain disadvantages too in decompositions of the foregoing kind. The units may not be totally self contained and trade offs between their efficiencies may be possible. For example a company could incur higher pumping costs in the sewerage stage in order to use larger treatment works and gain from sewage treatment economies of scale. Also indirect expenditure such as company headquarters expenses can be difficult to apportion to operating units below company level. (See Thanassoulis (forthcoming and 1997b) for further discussion of the foregoing points.)

## 4. CONCLUSION

This paper has given an introduction to the basic DEA models for assessing efficiency under constant and variable returns to scale. It has also outlined the use of DEA by OFWAT, the regulator of water companies in England and Wales, in the framework of its price determinations in 1994. DEA was used in a supporting role to OLS regression based estimates of comparative company operating efficiencies. Companies were given the best of the efficiency rating offered by DEA or OLS regression, so far as clean water operating expenditure is concerned. The efficiency ratings had an impact on the cap placed on the company's charges though the link between the efficiency rating and the cap determination is not a direct one. The potential impact of the DEA analyses is nevertheless very substantial in view of the very large charges involved aggregated both over the country and over several years. The paper concluded with certain unresolved generic issues affecting the use of DEA in regulation.

## REFERENCES


Banker, R. D.(1984), “Estimating Most Productive Scale Size Using Data Envelopment Analysis”, *European Journal of Operational Research*, 17, pp. 35-44.

Banker, R. D., Charnes, A. and Cooper, W.W., (1984) "Some models for estimating technical and scale inefficiencies in data envelopment analysis.*"*, *Management Science,* 30, 9, pp.1078-1092

Charnes, A., Cooper, W. and Rhodes, E., (1978) "Measuring the efficiency of decision making units", *European Journal of Operational Research* 2, pp. 429-444.

Banker R. D. And Thrall R. M. (1992) “Estimation of Returns to Scale  Using Data Envelopment Analysis”, *European Journal of Operational Research, Vol. 62, pp. 74 - 84*.

Farrell M. J. (1957) The Measurement of productive efficiency. *Journal of the Royal Statistical Society.*;120 pp. 253-290.

OFWAT (1993) “Research Paper No. 2 — Modelling Water Costs 1992-3”, Prepared by M. Stewart for OFWAT, 7 Hill Street, Birmingham B5 4UA, England.

OFWAT (1994a) “1993-94 Report on the Cost of Water Delivered and Sewage Collected”, OFWAT 7 Hill Street, Birmingham B5 4UA, England.

OFWAT (1994b) “Future Charges for Water and Sewerage Services: The outcome of the Periodic Review”, OFWAT 7 Hill Street, Birmingham B5 4UA, England.

OFWAT (1995), “South West Water Services Ltd”, HMSO.

Thanassoulis E. (1995), “Assessing Police Forces in England and Wales Using Data Envelopment Analysis”, *European Journal of Operational Research,* Vol. 87, pp. 641 - 657.

Thanassoulis, E. (1997a) Duality in Data Envelopment Analysis under Constant Returns to Scale, *IMA Journal of Mathematics Applied in Business and Industry, Vol. 8, No. 3 (July 1997),  pp. 253 - 266.*

E. Thanassoulis (1997b), Using DEA to Estimate Potential Cost Savings in Sewerage, *Working Paper No. 269,* Business School, Warwick University, Coventry CV4 7AL, UK.

E. Thanassoulis (forthcoming) The Use of Data Envelopment Analysis in the Regulation of UK Water Utilities: Water Distribution, *European Journal of Operational Research.*