Efficiency Measurement in the Privatised English and Welsh Water and Sewerage Industry 1992-2004.

Helen Lay

Indepen, London

&

Melvyn Weeks

Faculty of Economics

University of Cambridge

August 17, 2006

Paper Presented at the North American Productivity Workshop

Stern School of Business

New York University

New York, NY

June 27 - June 30, 2006

The regulation of the Water Industry in England and Wales

We consider the problem faced by a public sector regulator, Ofwat, in regulating the English and Welsh water and sewerage industry.

Cost and operating characteristics for each firm are used to determine the degree to which efficiency may be improved, and ultimately price limits.

Ofwat makes this judgement based on the following components:

1. the overall **scope** for improvements in efficiency
2. **comparative efficiency** analysis
3. the efficiency gap
4. the proportion of any efficiency gap that should be closed (catch-up)
5. the speed of catch-up

Challenges to Ofwat's approaches

The opportunities to review the regulator's decisions are as follows.

- Regulated companies can require referral to the Competition Commission of any licence modifications proposed by Ofwat (including price limits).
- Parliamentary committees and the National Audit Office scrutinise the activities of regulators; they have not looked in detail at the efficiency question in water.

Principal recommendation of Independent Steering Group (2004)

In view of the widespread industry dissatisfaction with Ofwat's efficiency methodology, and the large number of studies carried out in this area, we recommend instituting a joint industry review of efficiency studies and econometric approaches with the aim of developing a common industry approach for the 2009 review.

Objectives

In this study our primary focus will be on comparative efficiency

- We consider Ofwat's current approach to **identify** the efficiency frontier and estimating relative efficiencies
- We examine alternative model-based routes to identifying efficiencies, and in particular the separation of efficiency from heterogeneity.
- We will consider both Classical and Bayesian approaches
- Issues: dealing with a small number of comparators; the need for interval estimation of inefficiencies; the benchmark firm is unknown

$RPI + Y_i - X_i$ **Price Cap Regulation in the Water Industry**

The current approach adopted by the water regulator (Ofwat) represents a combination of both econometric outputs and qualitative decisions.

A point estimate of inefficiency, $\widehat{X}_i$, for firm is provided for the i^{th} firm. This is then subject to a number of **regulator interventions**.

1. adjustments for company specific factors
2. adjustments for future expenditure
3. the selection of a benchmark company
4. the determination of a catch-up factor based on a percentage of the distance of a given firm to the frontier (i.e. to the benchmark company). See Fig. 1.

Catch-up

After the benchmark company is selected, Ofwat calculates the distance of each company to the benchmark, i.e.

Ofwat then places the companies into the A to E bandings.

Ofwat further sub-divides the bandings into half-bands.

For each of these half bands, the mid-point of the band is determined to be the total potential for catch-up.

Ofwat assumes that a percentage of the gap between the most efficient and least efficient must be closed within the review period (60% for water and sewerage operating expenditure in PR04).

Table 1 presents the individual company catch-up factors for companies, conditional on the allocation of companies to bands A-E.

Overview

1. Price Cap Regulation in the water Industry
2. **Small** number of Comparators
3. Efficiency Not Directly Observable
4. Incorporating Regulator Uncertainty as to the Benchmark Firm
5. **Absolute** versus **Relative** Efficiency
6. **Point** versus **Interval** Estimation of Efficiencies
7. **Exact** versus **asymptotic** inference
8. Preliminary Analysis of Water Industry panel data

Frontier Cost and Production Function: Review

The empirical measurement of technical efficiency, $TE(y, \mathbf{x})$, requires definition of a transformation function.

Definition 1 *In general terms TE characterises the relationship between an observed cost (production) and a potential (cost production).* ***Single output*** *- formulate this in terms of TFP, the ratio of actual output to the optimal value as specified by a production function, $f(\mathbf{x})$, using input vector $\mathbf{x}$*

$$y \leq f(\mathbf{x})$$

An output based Debreu-Farrell style measure of technical efficiency is

$$TE(y, \mathbf{x}) = \frac{y}{f(\mathbf{x})} \leq 1 \quad (1)$$

$$y_i = f(\mathbf{x}_i, \boldsymbol{\beta}) TE_i \quad (2)$$

Frontier Cost and Production Function: cont.

Assuming that the production (or cost model) is linear in logs of the variables, then

$$\begin{aligned} \ln y_i &= \ln f(\mathbf{x}_i, \boldsymbol{\beta}) + \ln(TE_i) \\ &= \ln f(\mathbf{x}_i, \boldsymbol{\beta}) - u_i \\ u_i &= -\ln(TE_i) \end{aligned}$$

$u_i \geq 0$ is a measure of technical inefficiency

Remark 1 *The production frontier assumes that conditional on a frontier all points are either on the frontier or lie beneath. A cost frontier assumes the reverse such that we write*

$$\begin{aligned} \ln C_i &= \ln f(\mathbf{x}_i, \boldsymbol{\beta}) + \ln(TE_i) \\ &= \ln f(\mathbf{x}_i, \boldsymbol{\beta}) + u_i \end{aligned}$$

Inefficiency is not Directly Observable

Different representations of residual terms for a production (cost) frontier. Inefficiency is *negative* (*positive*)

$$\varepsilon_i = v_i \tag{3}$$

$$\varepsilon_i = v_i \pm u_i \tag{4}$$

$$\varepsilon_{it} = v_{it} \pm u_i \tag{5}$$

$$\varepsilon_{it} = v_{it} \pm u_{it} \tag{6}$$

v_i denotes a symmetric error term (perhaps normal); u_i is the non-negative inefficiency component.

Identification of Firm-Specific Efficiencies

Remark 2 *Assumptions re: the nature of firm level unobserved characteristics together with any distributional assumptions, serve to both identify firm-specific efficiencies, and dictate the nature of subsequent interpretation.*

Identification - OFWAT: Ad hoc adjustment of model estimated residuals for unmeasured firm-specific factors.

This type of intervention represents one way of accounting for unobserved-firm specific heterogeneity, and thereby identifying firm-specific efficiencies with cross-section data.

Identification of Firm-Specific Efficiencies: Panel Data

Panel data: RE or FE specification precludes the necessity for these ad hoc adjustments.

Remark 3 *Panel data affords separate identification of a firm-specific effect from a general stochastic error term*

FE *- the absence of distributional assumptions serves to identify* ***relative*** *efficiencies*

RE *- the imposition of the classic one-sided error distributions, facilitates identification of* ***absolute*** *efficiency measures.*

Identification of Firm-Specific Efficiencies: Panel Data cont

Problem with both specifications - confounding of the firm specific inefficiency component with other time invariant firm-specific effects which are observed.

These might include, for example, topography and regional cost differences providing they affect companies in a similar way year on year.

Regulator intervention might still required?

Separating the Inefficiency from Unobserved Heterogeneity

Remark 4 *identification of both inefficiency and heterogeneity using panel data is not possible in either the standard FE or RE specifications given that α_i and u_i cannot be separated.*

Following earlier work by Kumbhakar and Hjalmarsson (1995), Greene (2005) proposed extensions of both R and F effects models, with the intention of separating the inefficiency term from unobserved heterogeneity.

Solution: allow the inefficiency term to be time varying.

The FE variant becomes a SFM with the addition of dummy variables to mop up firm level differences which are distinct from firm level inefficiencies.

Measures of Efficiency

We consider a multi-input, multi-output production technology.

A firm uses a non-negative $K \times 1$ input vector $\mathbf{x} = (x_1, ..., x_K)'$ to produce a non-negative $M \times 1$ output vector $\mathbf{y} = (y_1, ..., y_M)'$.

$\mathbf{w} = (w_1, ..., w_K)'$ denotes a $K \times 1$ vector of input prices.

A Distance Function Approach

The production function is a natural representation of technology for firms with a **single** output.

Multiple output technology can be represented by a set of production functions under strict assumptions.

An alternative representation in quantity space is the **Distance Function** (DF)

An **input distance function** measures the degree to which the input vector can be proportionately contracted with output fixed.

Defining the Input Distance Function

Letting $F(\mathbf{y})$ denote the set of all input vectors, $\mathbf{x}$ that can produce the output vector $\mathbf{y}$, namely

$$F(\mathbf{y}) = \{\mathbf{x} : \mathbf{x} \textit{ produce at least } \mathbf{y}\} \tag{7}$$

the input DF

$$D(\mathbf{x}, \mathbf{y}) = \max\{\theta : \mathbf{x}/\theta \in F(\mathbf{y})\} \tag{8}$$

gives the maximum proportional reduction in inputs that is possible with the output vector held fixed.

Input Distance Function

Given $TE(\mathbf{y}, \mathbf{x}) = 1/D(\mathbf{y}, \mathbf{x})$ we may write

$$x^* D(\frac{x_2}{x^*}, ..., \frac{x_K}{x^*}, \mathbf{y}) TE = 1$$

where x^* denotes an arbitrarily chosen input; $\xi = x^{*-1}$,

Taking logs

$$\begin{aligned} 0 &= \ln(x^*) + \ln D(\frac{x_2}{x^*}, ..., \frac{x_K}{x^*}, \mathbf{y}) + v + \ln(\exp(-u)) \\ -\ln(x^*) &= \ln D(\frac{x_2}{x^*}, ..., \frac{x_K}{x^*}, \mathbf{y}) + (v - u) \end{aligned}$$

Characteristics of the Input Distance Function

1. $D(\mathbf{x}, \mathbf{y})$ is decreasing in y_m, $\;m = 1, .., M$
2. $D(\mathbf{x}, \mathbf{y})$ is decreasing in x_k, $\;k = 1, .., K$
3. $D(\mathbf{x}, \mathbf{y})$ is homogeneous of degree 1 in $\mathbf{x}$
4. $D(\mathbf{x}, \mathbf{y})$ is concave in $\mathbf{x}$

homogeneous of degree 1 in $\mathbf{x}$ implies

$$D(\xi\mathbf{x}, y) = \xi D(\mathbf{x}, \mathbf{y}), \quad \xi > 0.$$

A Distance Function Approach: Advantages

1. Approach does not require behavioural assumptions such as cost minimisation or profit maximisation.
2. the appeal of measuring **multiple output** inefficiencies in a primal setting is that the primary emphasis is on securing estimates of technical efficiency (TE)
3. TE is a physical notion which may be measured without reference to either **price information** or a specific **behavioural objective**.
4. no need to separate technical and allocative inefficiency
5. Coelli and Walding (2005) highlighted the use of a DF approach to the analysis of inefficiency in the public sector.

A Distance Function Approach: Application in the Water Industry

Given the role of the water regulator and the quinquennial price reviews, we are not able to maintain any assumptions related to the exogeneity of prices.

Input mix allocation efficiency measure the effects of deviation from least cost combination of inputs.

Example: a critical element in the input set, capital, is largely exogenous given that it is driven by population density.

Saal and Parker (2005) employ a quality adjusted multiple output DF to examine inefficiency in the UK water industry.

Panel Data Estimators of Inefficiencies

The standard panel data frontier model

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} - u_i + v_{it} \tag{9}$$

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} \tag{10}$$

$$\alpha_i = \alpha - u_i \tag{11}$$

i, t denote firms and time periods

y_{it} denotes the logarithm of output of firm i in period t

$\mathbf{x}_{it}$ is a $k \times 1$ vector of inputs, $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown parameters.

$v_{it} \sim (0, \sigma_v^2)$ an i.i.d. stochastic error term. $u_i \geq 0$ are measures of technical inefficiency which we initially assume are time invariant.

technical efficiency is given by $r_i = e^{-u_i}$.

Fixed versus Random Effects

Advantage: fixed effects specifications - the lack of distributional assumptions placed on the inefficiency terms.

Disadvantage: the estimation of inefficiency is *relative* to the most efficient firm.

Relative inefficiency of the i^{th} firm, u_i^*, is written as

$$u_i^* = \alpha_{(N)} - \alpha_i \tag{12}$$

$$r_i^* = e^{-u_i^*} \tag{13}$$

An estimate of technical inefficiency (Schmidt and Sickles (1984))

$$\widehat{\alpha} = \max(\widehat{\alpha}_i), \quad i = 1, ..., N \tag{14}$$

$$\widehat{u_i}^* = \widehat{\alpha} - \widehat{\alpha}_i.$$

Fixed versus Random Effects - cont

Fixed effect estimator: disadvantages:

1. the impact of observed time invariant firm specific effects cannot be determined.
2. the use of the max operator makes classical inference for u_i^* difficult. MCB - a technique which generates CIs for differences in parameter values between all populations and the best, represents a possible solution here (See Horace and Schmidt (1996)).
3. since the most efficient firm is unknown we **estimate**. Therefore firms with the smallest $\widehat{\alpha}$ may not be the most efficient.

If an assumption of the exogeneity of $\mathbf{x}_{it}$ is valid then a random effects specification for the inefficiencies brings some benefit.

Bayesian Fixed and Random Effects Stochastic Frontier Models

Bayesian - all effects are random; the classical fixed versus RE dichotomy is not relevant.

Key distinction - the nature of the marginal prior links between models - specifically between hierarchical and non-hierarchical models.

Koop Osiewalski & Steel (1997) [KOS] consider two variants of the Bayesian F model, both maintaining that firm-specific effects are marginally independent.

Bayesian Fixed and Random Effects Stochastic Frontier Models: Fixed Effects

The standard individual effects (SIE) variant places an improper uniform prior on α_i - this model is the counterpart to the classical FE model.

Firm efficiencies are again estimated relative to the most efficient firm BUT there is no assumption that the most efficient firm is known.

Given parameter uncertainty, the firm with the largest $\widehat{\alpha}$ (and smallest u_i) may not be the most efficient.

Solution - calculate the distribution of inefficiency of firm i, relative to j, $\forall j$, and weight by the probability that j is the most efficient.

The marginal posterior for Relative Inefficiency

The marginal posterior for relative inefficiency is a mixture distribution.

The probability mass at full efficiency

$$\Pr(\alpha_i = \max_j(\alpha_j)|y, \mathbf{x}) = \Pr(u_i^* = 0|y, \mathbf{x}) \equiv P_i \tag{15}$$

The density for $u_i^* > 0$ is

$$\begin{aligned} p(u_i^*|y, \mathbf{x}) &= \sum_{j=1, j\neq i} p(u_i^*|y, \mathbf{x}, u_j^* = 0) P(u_j^* = 0|y, \mathbf{x}) \qquad (16) \\ &= \sum_{j=1, j\neq i} P_j p(u_i^*|y, \mathbf{x}, u_j^* = 0) \qquad (17) \end{aligned}$$

Bayesian Random Effects

The Bayesian counterpart to the RE model imposes prior links between individual effects by assuming different values for α_i, but from the same distribution.

Point of departure: firm-level inefficiencies are distributed i.i.d exp.

$$p(u_i|\lambda^{-1}) = \lambda^{-1}\exp(\lambda^{-1}u_i) = f_G(u_i|\lambda, 2)$$

with mean λ.

Different variants of the random effects model are based upon different treatments of the parameter λ.

Assume prior links between individual effects:

1. effects drawn from a common efficiency distribution

2. means of effects can be related to certain firm characteristics $\Longrightarrow$ Varying Efficiency Distribution

Bayesian Random Effects - cont

A simple Bayesian RE model is given by

$$\begin{aligned} y_{it} &= \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} - u_i + v_{it} \\ p(v_{it}) &= N(0, \sigma_v^2) \\ p(\sigma_v) &= IG(s, M) \propto exp(-s/2\sigma_v^2)(\sigma_v^2)^{-\frac{(M+1)}{2}} \\ p(u_i) &= \lambda_i^{-1} \exp(u_i/\lambda_i) \\ p(\boldsymbol{\beta}) &= N(0, \zeta), p(\alpha) = N(0, \zeta) \end{aligned}$$

inefficiency u_i is specified as exponentially distributed with mean λ.

Specification requires elicitation of a hyperparameter or the use of a hierarchical prior.

Bayesian Random Effects - cont

Use of a **hierarchical prior** treats λ as a random variable.

Sub-variants again depend upon the strength of prior links.

General model parameterises the mean parameter giving $\lambda = \exp(-\mathbf{z}_i'\boldsymbol{\gamma})$, $\mathbf{z}_i = (z_{i1}, z_{i2}, ..., z_{im}')$ is a $m \times 1$ vector of binary characteristics, $\boldsymbol{\gamma}$ is a conformable vector of unknown parameters.

z_{i1} is the unit vector.

Reparameterising the mean parameter as

$$\exp(\mathbf{z}_i'\boldsymbol{\gamma}) = \prod \kappa_j^{z_{ij}},$$

$\kappa_j = \exp(\gamma_j)$ are i.i.d Gamma distributed with parameters a_j and $b_{j.}$

Restricted model: $m = 1$, $\mathbf{z}_i = z_{i1}$

Bayesian Random Effects - cont

General and restricted models:

- Varying Efficiency Distribution (VED)
- Common Efficiency Distribution (CED).

CED imposes stronger prior links.

VED allows mean inefficiencies to vary dependent upon the characteristics of the firm, the restricted model assumes an i.i.d structure.

Bayesian Random Effects: CED

Characteristic of the VED - allows for a departure from the standard RE assumption that the firm-specific effects are independent of the regressors.

Remark 5 *Within a classical setting Chamberlain (1980, 1982) introduced the correlated RE model which by parameterising the random effects in terms of one or more regressors circumvents this strong assumption.*

the classical model is operationalised for both continuous and discrete components of $\mathbf{z}_i$*. In the Bayesian context, the restriction that* $\mathbf{z}_i$ *are all binary variables significantly simplifies the computations, given that the* κ_j*'s have a simple Gamma distribution.*

An Bayesian Extended Random Effects Model

Consider a random firm effect $\alpha_i = \alpha + w_i$

w_i is a two-sided random component.

u_{it} denotes time-varying firm-level inefficiency

$$
\begin{aligned}
\ln y_{it} &= \alpha_i + \beta^T \mathbf{x}_{it} + v_{it} + w_i - u_{it} \\
p(v_{it}) &= N[0, \sigma^2], p(\sigma) = IG(s, M) \\
p(w_i) &= N[0, \tau^2], p(\tau) = IG(r, T) \\
p(u_{it}) &= (1/\lambda) \exp(-u_{it}/\lambda), p(\lambda) \\
p(\beta) &= \propto 1.
\end{aligned}
$$

λ - use hyperparameter or hierarchical prior

Bayesian Random Effects Stochastic Frontier Models

Assume prior links between individual effects:

1. effects drawn from a common efficiency distribution (CED model)

2. means of effects can be related to certain firm characteristics $\Longrightarrow$ Varying Efficiency Distribution (VED)

Standard Panel RE has Hierarchical prior:

$\alpha \sim N(\mu_\alpha, \sigma_\alpha^2)$

$\mu_\alpha, \sigma_\alpha^{-2}$ have a Normal-Gamma prior distribution.

In this particular context we require a Hierarchical prior which reflects the non-ive property of u_i

Remark 6 *Placing an informative prior for u_i allows us to distinguish u_i from α (*Relative v. Absolute efficiency*)*

An Unbalanced Panel of Water Companies

1989 Water Act enabled parts of the UK water industry to be passed to the private sector.

The Act created the National Rivers Authority and the 10 Water and Sewerage Companies (WaSCs).

Existing constraints (financing and dividend policy) on 29 small, privately owned water-only companies (WoCs) were removed, transforming them into normal public ltd. liability companies.

The 10 WaSCs were protected from takeover for 5 years.

The smaller water-supply only companies were subject of takeovers straightaway.

The number of WoCs has fallen from 29 in 1989 to the current 12.

An Unbalanced Panel of Water Companies: 1992 to 2002

The dataset used in this study consists of an unbalanced panel for water and sewerage companies and water only companies.

The evolution of firms is driven by the process of mergers and acquisitions that occurred between 1995 and 2002.

When mergers took place we have considered the merged entity as a new firm entering the panel.

Example: merger in 1996 between East Surrey Water and Sutton and District Water, the new firm Sutton and East Surrey Water was created.

Other mergers are illustrated in Table 2.

Data: Outputs

Main source of output and input data - annual company June Returns.

Following Stone and Webster (2004) and Saal and Parker (2005) we consider two specific **outputs**.

i. **water delivered** is the estimated total volume of water supplied to the boundary of each customer's property and includes leakage.

ii. **connections to the water network**

a distinct output given that input requirements are likely to be different.

Data: Inputs

Inputs - we include measures of employment, power, materials, and other input costs

All costs, except power, are rebased to 1992-93 prices using the all-items RPI.

Power costs are similarly rebased using the average level of the PPI for electricity costs over the relevant financial year.

All inputs denote the sum of those incurred in *water resources and treatment* plus those incurred in *water distribution*.

Capital costs: gross replacement cost of capital employed in "water service provision" at 31st March.

Capital employed in sewerage service provision is reported separately in the Annual Returns.

Including Observed Heterogeneity

How do we introduce observed heterogeneity (z)?

Kumbhakar and Lovell (2000) - very little theoretical guidance

1. as a component of production technology i.e. $\mathbf{z}$ directly effect input requirements.
2. a determinant of productive efficiency i.e. parameterise the inefficiency error component using $\mathbf{z}$
3. both

We do 3.

1. Network density measured as total population served per kilometre of water mains represents a partial control for exogenous cost differentials due to population density.

2. We utilise a binary indicator of whether firm is a WASC/WoC.

Results: Unit Cost Comparisons

Table 2 show how unit costs have changed since 1992-93.

Both the water and sewerage unit costs of operation have reduced since 1992-93, reflecting the falling trend in operating costs.

In recent years unit costs of operation have remained broadly unchanged as increases in the volumes supplied to customers have been matched by similar increases in operating costs arising from new quality obligations.

Tables 3 presents changes in unit costs of operation per property billed between 1995-96 and 2001-02.

In many cases **water service** unit costs of operations per property billed have fallen since 1995-96

Example: Southern Water has shown a reduction of 34%, giving it the second lowest unit costs of operations in the industry.

Model Comparisons

Classical

1. Corrected OLS: Cobb-Douglas
2. Corrected OLS: Translog
3. SFA (exponential): Pooled, Translog
4. Random Effects (exponential): Translog

Bayesian

1. Random Effects: Common Efficiency Distribution
2. Random Effects: Varying Efficiency Distribution

Results Presented

For each model we focus on measures of firm-level efficiency.

Classical estimators: report point estimates without any measures of parameter uncertainty.

Bayesian: report the 95% posterior credible interval for efficiency estimates.

We also present similar intervals for the ranks

Priors From Ofwat's Models

Ofwat's relative efficiency assessment for 2001-02

- Yorkshire Water used as the benchmark company for delivering water services
- Thames Water is the most efficient company for delivering sewerage services
- These findings are based upon a combination of an econometric model and qualitative adjustment for special factors.
- Thames was not used as the benchmark company since it had a large adjustment factor applied

Commentary on Results: Classical

Relative to Unit Costs Comparisons Inference on water company relative efficiencies is partially consistent with the story on unit cost comparisons (at the extremes)

Classical Inference: pooled models Pooled COLS and pooled SFA show similar results in terms of rankings. Little time variation in efficiencies

Most efficient companies: Portsmouth and Cambridge

Most inefficient companies: South East and Yorkshire

Classical Inference: Random Effects Random effects model with time invariant efficiency shows Portsmouth and South East water as the most (least) efficient respectively

Commentary on Results: Classical - cont

Classical Inference: True Random Effects We observe a number of differences in terms of the profile of time varying efficiencies.

This model confirms the regulators choice of Thames as the most efficient company in 2001.

Ofwat Approach: econometric model with regulator intervention

This approach: using parametric assumptions to separately identify efficiency and firm-specific heterogeneity

Commentary on Results: Bayesian

Varying Efficiency Distribution: exponential Treating the mean of the inefficiency distribution as a random variable, and paramterised using a linear index: $WASC_D\gamma_1 + Merger_D\gamma_2$

Common Efficiency Distribution: exponential A restricted variant of the VED model. Impose stronger prior links by assuming a constant mean.

Although VED indicated a better fit based on the Deviance Information Criteria (DIC), the results are quite similar in terms of posterior distribution of efficiencies and rankings.

Further Work

Adjust water volumes by water quality

Leakage: Undesirable by-products of the production process.

Leakage varies significantly across water companies and over time. Fernandez, Koop and Steel (1994) have proposed methods to deal with undesirable by-products of the production process

Extended Bayesian Random Effects Model

Separating Firm-level efficiencies and unobserved heterogeneity