

# **Assessing Ofwat's Efficiency Econometrics**

A report for Water UK, March, 2004

Professor John Cubbin  
City University  
[j.s.cubbin@city.ac.uk](mailto:j.s.cubbin@city.ac.uk)

# Assessing Ofwat's Efficiency Econometrics

## TABLE OF CONTENTS

Executive summary .....	iii
1. Introduction.....	1
2. Methodology .....	2
2.1 <i>The sources of error in econometric models</i> .....	3
2.2 <i>Collecting and assessing evidence on the quality of the models</i> ...	5
3. Evaluation of each model .....	14
3.1 <i>Operating cost models: Water</i> .....	15
3.2 <i>Operating cost models: Sewerage</i> .....	21
3.1 <i>Capital Expenditure models: Water</i> .....	29
3.2 <i>Capital Expenditure models: Sewerage</i> .....	35
4. Aggregation Issues .....	41
4.1 <i>Problems of Aggregation</i> .....	41
4.2 <i>Summary of Findings</i> .....	47
5. Limitations and potential questions.....	47
6. The consequences of error: a simulation .....	53
7. Summary and implications.....	56
Appendix A1: The econometric framework.....	57
Appendix A2: Measures of goodness of fit used in the Tables .....	61
Appendix A3: Sewerage non-infrastructure model .....	63

# Assessing Ofwat's Efficiency Econometrics

A report for Water UK, March, 2004

## Executive Summary

This paper reports on the results of estimating the percentage of the residual from Ofwat's comparative efficiency models "which can be attributed to inefficiency as opposed to random noise, uncertainties and omitted explanatory factors."

The method involves looking at each model in turn, looking at the evidence for the presence of the following contributions to model error:

- A. Sampling error
- B. Error in the measurement of the dependent variable
- C. Variables excluded from the analysis
- D. The explanatory factors being imperfect proxies for what they are meant to measure (errors in variables)
- E. The approximation inherent in the need to use a particular mathematical form

The assessment involves a consideration of factors such as the engineering and economic knowledge of the industry, the number of observations used, and the plausibility of the adjusted cost ratios implied by Ofwat's analysis, and evidence of outliers.

Having taken account of the argument that some of the measurement error may cancel out across the models, our estimates of the proportion of the residual that may be due to efficiency variation is shown as Table ES1.

**Table ES1 Summary of findings (rounded to nearest 5%)**

	Estimated proportion of residual attributable to efficiency (point estimates)
<b>Operating cost</b>	
<i>Water</i>	40%
<i>Sewerage</i>	50%
<b>Capital maintenance</b>	
<i>Water</i>	25%
<i>Sewerage</i>	325%

These proportions take no account of errors in the models due to trade offs between opex and capital maintenance, which will create artificially low efficiency frontiers. We then consider the implications for the accuracy of Ofwat's classification into efficiency groupings.

We find that, for the range of standard error of estimates found by Ofwat, sampling error alone will give rise to significant amounts of misclassification of companies.

**Table ES2: sampling error, other errors and banding misclassification**

(1) Row number	(2) % of variance due to inefficiency	(3) Residual range (adjusted cost ratio)	(4) % of companies Misclassified by 1 or more bands	(5) % of companies Misclassified by 2 or more bands
1	100%	1.04	0%	0%
2	100%	2.13	31%	2%
3	90%	1.93	41%	4%
4	80%	1.88	47%	6%
5	50%	2.13	62%	20%
6	40%	2.15	66%	22%
7	35%	2.15	68%	23%
8	25%	2.13	73%	27%

Table ES2 (a condensation of Table 6.1) shows the following:

1. When sampling error is the only error and the efficiency range (**the ratio of highest to lowest cost company**) is very small (**1.04**) companies are correctly classified into efficiency bands A to E. (Row 1)
2. However when the efficiency range increases to a more typical level of 2.1, then 30% of companies are misclassified. (Row 2)
3. If the percentage of the residual which represents inefficiency approaches the values Ofwat suggests (80 and 90%) then there is substantial misclassification. 41% to 47% of companies are misclassified by at least one band and 4 to 6% by two or more bands. (Rows 3 and 4)
4. Once other sources of error are introduced the average rate of misclassification rises to over half, with 20% or more of companies misclassified by more than one class. (Rows 5-8)

This suggests that, in applying the findings of the econometric efficiency analysis Ofwat needs to take due account of the likelihood that a substantial fraction of the companies will be assigned to the wrong category.

# Assessing Ofwat's Efficiency Econometrics

A report for Water UK, March, 2004

Professor John Cubbin  
City University  
j.s.cubbin@city.ac.uk

## 1. Introduction

Efficiency analysis forms a key part of Ofwat's price determination process. Its results can affect the distribution of revenue between companies, and ultimately the attractiveness of investment in those companies. It can affect how hard it is for companies to make the savings necessary to meet or exceed the efficiency targets.

Done correctly it can ensure fairness across companies and their customers. Conversely, a poor or inappropriately applied analysis can have consequences for the whole industry. If it is seen as unduly harsh or arbitrary it may make the raising of capital only possible on the promise of higher returns: lower prices now may mean higher prices later on. In the extreme case it could undermine the credibility of the whole regulatory framework, with severe long term consequences. On the other hand, failure to take adequate account of the potential for cost reduction can give companies and their shareholders easy windfall gains which could undermine the credibility of the regulatory framework at a political level.

It is clearly in the industry and the country's interest if a consensus can be arrived at regarding the precise weight to which the econometric results can bear. It is true that companies in the industry face what may be regarded as a zero-sum game. Each model will have companies which are rated highly (and the opposite.) Changing the model or reducing the weight is likely to impact negatively on the first and positively on the second. But it is in the interests of the industry as a whole to arrive at as accurate an assessment as possible.

Ofwat claims that its models undergo a rigorous feedback process designed to produce the very best models attainable with the available data. Ofwat works with the industry, taking on board comments and trying out modelling suggestions made by the companies and others. Ofwat claims that the statistical modelling is robust, and that "there may be *up to* a 40% difference between the most and least efficient companies. However *some* of the difference may result from the limitations of the data."<sup>1</sup> (Emphasis added.)

How much is "some"? The key question I was asked to address is:

"... an estimate of the percentage of the residual from each model which can be attributed to inefficiency as opposed to random noise, uncertainties and omitted explanatory factors."

---

<sup>1</sup> Ofwat *Water and Sewerage service unit costs and relative efficiency: 2001-2 Report.*

This is a challenging assignment which could be approached in a number of ways, for example:

1. (Top down, performance-based, *ex post*) examine the past history of cost movements to estimate the degree of convergence of comparable costs and the degree to which they converge to common values
2. (Bottom up, *ex ante*) prior examine each equation in turn and assess its likely errors in terms of “random noise, uncertainties and omitted explanatory factors”
3. A purely statistical approach, looking at the extent to which there are common company efficiency effects across the models, or examining the decomposition of the residuals into symmetric unbounded error terms and asymmetric, bounded or efficiency terms.

The first would require a major modelling effort which is beyond the scope of the present exercise. The third involves assumptions about the nature of managerial efficiency and is not reported here.

This report therefore focuses on the second of these. As part of our framework, we initially used as a first stage the criteria proposed by Ofwat's econometric adviser, Professor Mark Stewart:

- 1 Consistency with technical knowledge of the industry (engineering & economic)
- 2 Statistical validity: consistency with the data
- 3 Suitability for purpose: use for relative efficiency measurement
- 4 Simplicity
- 5 Validity of available data on variables used in model

However, only some of these are directly relevant to the central question, which requires a consideration of the nature of econometric models and how they can fail to be 100% accurate. We therefore developed an approach which considers systematically the sources of error in the models which could undermine their ability to measure relative efficiency.

This is set out in detail in Appendix A1 and summarised in Section 2 below, which also sets out our methodological approach.

To anticipate our results we reproduce Table 4.5 (rounded to the nearest 5%) below which gives a summary of the point estimates we derive. It is important that these be considered in conjunction with the qualifications and not be quoted out of context.

**Table 4.5 Summary of findings (rounded to nearest 5%)**

	<b>Estimated proportion of residual attributable to efficiency (point estimates)</b>
<b>Operating cost</b>	
<i>Water</i>	40%
<i>Sewerage</i>	50%
<b>Capital maintenance</b>	
<i>Water</i>	25%
<i>Sewerage</i>	35%

The paper is structured as follows. Section 2 describes the methodology adopted. section 3 provides an evaluation of each model. Section 4 addresses issues of aggregating the individual models to an overall assessment, and in particular the “swings and roundabouts” and “double jeopardy” arguments. Section 5 looks at the limitations of the methods and some of the potential criticisms of the approach. In section 6 we examine how the findings might affect the use that Ofwat should make of its efficiency modelling. This is a complex issue, but we attempt to shed light on it with the aid of some results from simulations of the data generation and estimation processes. Finally section 7 presents a summary and some implications.

## **2. Methodology**

In this section we look first at the reasons why the residuals in econometric models of cost may reflect influences other than the relative efficiency of the companies in the model.

Section 2.2 then describes how we assess the evidence on the properties of Ofwat’s models, and Section 2.3 describes and explains the reporting framework for each model.

### **2.1 The sources of error in econometric models**

Appendix A1 sets out in some detail the general econometric framework and identifies the following sources of error in econometric models. These are summarised below:

- A. Even with totally accurate data and models, estimates are just estimates and subject to sampling error as long as there are limited observations
- B. There may be error in the measurement of the dependent variable
- C. There may be variables excluded from the analysis
- D. The explanatory factors may themselves be proxies and/or subject to errors of measurement.
- E. The wrong mathematical form may have been chosen to approximate the relationship between costs and their drivers.

Professor Mark Stewart has advised Ofwat on many aspects of their econometric estimation of efficiency, and has devised five criteria for evaluating the models. How do these criteria relate to the approach we are using?

1 Consistency with technical knowledge of the industry (engineering & economic)

In so far as the model does not represent industry knowledge this could have an impact on errors C, D, and E.

2 Statistical validity: consistency with the data

A poor fit in terms of a large standard error of estimate or, equivalently, a wide range of implied inefficiencies leads to larger sampling errors (error A). It may also be indirect evidence of errors B, C, D, and E.

3 Suitability for purpose: usefulness for relative efficiency measurement.

This criterion is not directly relevant to an estimate of the inefficiency component of the residual. There are two kinds of issue here. First, would the application of the model in the usual way create perverse incentives? Second, is the model so bad on other grounds that it would be totally useless, even given a discount factor of 90% or more!

Ofwat has chosen not to estimate some models it considers on an *ex ante* basis to have bad incentive properties. This may have resulted in the rejection of models that are superior on other grounds. An alternative approach would have been to estimate the best model but apply it in a way that avoided poor incentives.

Suitability for purpose should be seen as a hurdle rather than a graduated assessment. A model must pass this test in order to be considered valid.

4 Simplicity.

Again this criterion is not directly relevant to the assessment of the error component. In econometric analysis it is a convention, based on the principle of Occam's razor, to prefer the simpler of two models with the same statistical properties.

Simplicity is at a particular premium when there are few observations, as is often the case in water industry econometrics. However, a model whose simplicity is inconsistent with industry knowledge leads to models that omit variables and have a higher sampling variance, leading to equation error of type C.

This may also lead an increased use of proxy measures in order to use a single explainer to simulate the effect of more than one factor. For example, population may, as a scale factor, approximate both mains length and number of connections, but it is an imperfect measure of each. The resulting measurement error in the explanatory variable is a type D error.

5 Validity of available data on variables used in model.

This criterion relates directly to errors B and D.

The above discussion is summarised in Table 2.1. The letter Y in a box indicates where an equation meets the Stewart criterion provides evidence about the likely size of a particular source of equation error.

**Table 2.1: The relationship between Stewart criteria and equation errors**

	Sources of equation error				
	A: Sampling variation	B: Measurement Error (dependent variable)	C: Omitted variables	D: Proxies or errors in measurement	E: Mathematical form
<b>Stewart Criterion</b>					
1 Consistency with technical knowledge			Y	Y	Y
2 Statistical validity: consistency with the data	Y	Y	Y	Y	Y
3 Suitability for purpose: use for relative efficiency measurement					
4 Simplicity			Y	Y	
5 Validity of available data on variables used in model		Y		Y	

## **2.2 Collecting and assessing evidence on the quality of the models**

Companies belonging to Water UK were invited to send me comments on each model. I have received comments from 9 companies, representing a range of views. (As may be expected, most but not all of the comments were from WASCs.) In some cases every model was commented on; in other cases attention was focussed on matters of particular concern. I should like to thank all those who provided me with their views. However I have treated these submissions with a some scepticism, given companies may have vested interests, and have formed my own judgements.

As an aid to judgement I built a statistical model simulating the impact of different breakdowns of the overall error into its components. Examining several parameter combinations over many simulations of regression estimates gave an indication of how different types of error impacted on cost ratios, standard errors and other diagnostic statistics. The simulation model was later used to assess the impact of different error combinations on the accuracy of Ofwat's classification into efficiency bands (reported in section 7). In addition, over the last 11 years I have had discussions with several companies across the range of efficiency bands concerning Ofwat's models. For reasons of commercial confidentiality I have preserved the anonymity of the companies providing comments.

Company comments were concerned primarily with views about consistency with industry knowledge and data reliability, but also contained information about, for example, the implied range of efficiency scores – data which were not always available during the early phases of the project.

I have also referred to a number of other documents assessing the quality of Ofwat's equations. Ofwat's own reports and other communications with companies are obviously very helpful in describing and explaining the background to the models. The Nera report on the capital maintenance models, the Competition Commission reports on East Surrey and Mid-Kent Water, and Ofwat's own publications were also helpful in providing views from different perspectives.

In the end, however, I have been asked to provide as objective an assessment as possible, based on judgement developed over three decades of experience trying to estimate practical and meaningful econometric models.

*A summary of the Competition Commission's views*

The Competition Commission (and its predecessor, the Monopolies and Mergers Commission) has had occasion to express its views on Ofwat's econometrics both in merger cases and in references arising from companies' non-acceptance of Ofwat's price control proposals.

Although the particular panel members may be different in different cases there is a reasonable degree of consistency in the Commission's views:

1. Comparative efficiency assessment is an important aspect of Ofwat's price setting process and the loss of a comparator should not be accepted lightly.
2. Ofwat's overall methodology of econometric estimation is basically sound.<sup>2</sup>
3. However, there are questions about particular models, in particular those relating to capital maintenance.<sup>3</sup>
4. Ofwat should look at alternative methodologies, including stochastic frontier analysis and panel data approaches.<sup>4</sup>

At no point, however, does the Competition Commission venture a clear view on the proportion of the residual which is properly attributable to inefficiency.

*A summary of the views expressed in the Nera paper*

The Nera paper<sup>5</sup> makes comments on each capital maintenance model individually, but all the models have one feature in common which undermines their usefulness and

---

<sup>2</sup> Competition Commission : "Sutton and East Surrey Water Plc: A report on the references under sections 12 and 14 of the Water Industry Act 1991"

<sup>3</sup> See previous reference page 254. "We consider the capital maintenance modelling is likely to be less robust than the opex modelling..."

<sup>4</sup> Competition Commission inquiry "Vivendi Water UK PLC and First Aqua (JVCo) Limited: A report on the proposed merger" (also referred to in 2001-2 unit cost and efficiency report)

reduces the proportion of the residual which can properly be considered to be a relative efficiency index.

These are the excluded variables of a) the age of the asset, b) its condition, c) its utilisation and d) company policies about replacement and refurbishment.

Although the details vary, brand new assets may require a bedding-in period where additional expenditure is required. There should be a period where operating costs are reduced as well as maintenance, which can be done on a routine basis. As the asset ages, however, wear and tear typically results in the need for refurbishment or partial renewal, or complete renewal. Renewal is economic when the costs of the renewal fall below the present value of additional refurbishment and delayed renewal. How the balance falls between these two is partly a matter of judgment and partly a matter of each company's comparative advantage in the two types of activity.

The condition of an asset may also vary in ways not directly related to age. Different vintages are built differently, and may be subject to differing degrees of adversity in operating conditions. Some may be worked harder than others.

Additionally, there may be an interaction with capital enhancement programmes, in which there is significant difference between companies. The maintenance requirements of companies with large enhancement programmes is likely to be different from companies that have are simply maintaining existing assets. This is particularly true in the areas of sewage treatment and sludge disposal.

#### *Ofwat's own views*

Ofwat has responded to comments on its comparative efficiency analysis in a consultation document<sup>6</sup>. We summarise the conclusions below by extensive quotation.

1. Data quality: "We do accept that data quality limits the usefulness of any statistical analysis. We consider that our integrated approach has reduced this variability to a minimum so that this issue has only a marginal impact on the analysis and the use we put it too [sic]"
2. Company specific accounting policies: "We are checking how companies allocate leakage control costs and we will make adjustments to modelled operating and capital maintenance before deriving our relative efficiency rankings"
3. Variables in models lacking operational logic: "The form of this relationship must demonstrate engineering, operational, and economic coherence....  
"This particular model is trying to explain distribution operating costs and we find that these are driven not by engineering considerations but by the economics of serving customers... But, because the number of rural customers is a small proportion

---

<sup>5</sup> An investigation into the robustness of Ofwat's comparative efficiency analysis of capital maintenance expenditure. A report for Water UK, June, 1999

<sup>6</sup> *Setting Water and Sewerage Price Limits for 2005-10: Framework and Approach* 27 March 2003. available from Ofwat web site at:  
[http://www.ofwat.gov.uk/aptrix/ofwat/publish.nsf/Content/pr04\\_methodology\\_270303](http://www.ofwat.gov.uk/aptrix/ofwat/publish.nsf/Content/pr04_methodology_270303)

of the total we can safely ignore their influence on costs, and we can get a good measure of the urbanisation impact from the quantity of large diameter mains.”

4. Lack of stability in the econometric models over time: “Changes in cost structure may manifest as changes over time in the effect of explanatory variables in the econometric models. Since we can include only the most important explanatory variables in our econometric analysis there will be changes over time in the explanatory variables which appear in the models.

5. Quality of service or compliance: “Generally the quality of service and compliance is high for all companies....Of course a company exhibiting a relatively good or unusual service may seek to quantify the additional costs in providing this service level.”

6. No account is taken of asset inheritance. “There are no objective and consistent measures of asset inheritance. This means that it is not possible to take this into account in the econometric analysis. “

7. No account taken of economic choices made between operating and capital solutions: “At the moment our research tells us that business decisions are not contingent on the interplay with regulatory incentives and that they stem from the same thought and analysis processes that are extant in the competitive market.”

8. Too low explanatory power in the models. “In our view the explanatory power of the models is not low....

“We may not be able to explain all the variations in cost in an econometric model. We are limited statistically by the number of explanatory variables we can include in our models...There are other explanatory variable which may explain differences in costs but we have found that these are either impossible to measure or difficult to do so on a consistent basis.”

9. No attempt is made to account for likely errors. “There will be errors relating to:

- the quality of the data provided by the companies
- the variation in the way companies allocate their cost data
- the number of data sets available for consideration
- explanatory factors not included in the modelling

10. Ofwat then makes the point that data quality and cost allocation problems have improved over time.

“The number of independent data sets for the water analysis has declined and the scale of errors will probably have increased, but still not to the level of those arising from having to model only the ten sewerage companies. We estimate these errors are twice as much for sewerage as for water and we will adjust the model residuals accordingly - by factors of 20% and 10% respectively. On the final point, we deal with this through our process of considering special factors. Unfortunately this provides an asymmetric error adjustment as companies do not bring “beneficial” cost drivers to our attention. Nevertheless, our restrictions on company size for frontier setting purposes do stop a very small company with an especially benign working environment introducing errors into this aspect of our work.

11. "We are keenly aware of the impact of measurement and systematic errors in our work. We listen to expert advice on these and our regulatory judgements include allowances where necessary."

Although Ofwat makes some rebuttals, there is evidently a good deal of common ground in principle between Ofwat and its critics, and Ofwat declares itself willing to listen and to modify its conclusions where necessary.

On other issues, such as the "double jeopardy" argument concerning the setting of separate opex and capex frontiers, Ofwat is silent or impenetrable. They say "We recognise this is a serious issue" but appear to have made no attempt to modify their approach to deal with it.

No information is provided about how the figures of 20% and 10% discount figures are arrived at. (Unless it is that 10 observations is only half as good as 20 – or possibly it is based on the stochastic frontier modelling which has apparently been carried out.)

In the following section we report on a framework for the purpose of making assessment of the quality of each individual model and the assessment as a whole. As we shall see the number of observations is only one of several aspects of the analysis which needs to be taken into account.

### **2.3 Description of analytical and reporting framework**

The framework we use is designed to establish how much of the residual can properly be attributed to inefficiency.

In order to establish this we first try to answer the question for each model in turn. We then consider how the individual models add up to form an efficiency score for each service and broad category (opex/capex.) Whilst there are important issues concerning opex/capex trade offs, these are beyond the scope of the present paper.

For the individual models we adopt a standard approach, which can be made clearer by going through each category for a particular example. The *italics* are explanatory comments. Put simply, we know that the residual contains many elements besides pure relative efficiency

**Activity modelled:** Water Resources and Treatment *Title for model*

**Dependent variable:** Expenditure per property *Variable being modelled*

**Scale and/or explanatory factors:** Weighted industry average *Explanatory factors used, including any scale variables*

**Interpretation of mathematical properties:** Required expenditure on resources and treatment capital maintenance is the same per connected property for every company. *Some models are mathematically more complex than others and give rise to more or less issues for interpretation. Examples are provided where this is pertinent to evaluating the model.*

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost	Range excluding companies < 2.5% industry turnover
Simple unit cost	n.a.	n.a.	48.79	35.02	8.11

*In regression analysis the SEE (Standard Error of Estimate) gives an indication of the variability in the dependent variable after the effect of the explanatory factors has been taken. The usefulness of this depends on how the dependent variable is defined.*

*R<sup>2</sup>, which measures how much of the original variation has been explained, is even more dependent on the details of the equation. For these reasons we generally consider that the standard deviation of the residual (measured as a percentage of the fitted value) gives a better indication of the unexplained variation, and one which is independent of the model details. This is supplemented by the “range of adjusted cost” which is described in Appendix A2. Essentially it shows the multiple of costs adjusted for scale and explanatory factors of the company that is assessed by Ofwat as least efficient in the model compared with the company assessed by Ofwat in its model as most efficient. A figure 2.00 here would imply a potential for savings of 50% for the highest-cost company.*

*Taken together these give some light on the plausibility of the data and model, and indicate the likely sampling error to be encountered when applying it with the given number of observations.*

**Comments from industry knowledge:** This model is too simple to measure the costs for a complex and diverse set of processes. There are probably insufficient observations to arrive at a satisfactory model. The number of properties is only indirectly associated with higher resource and treatment costs.

*This section is for comments on the model's applicability given the likely true cost drivers.*

*The next section applies the information already considered to look in turn at the major reasons why a model may not accurately reflect relative efficiency.*

**Sources of error:**

**Omitted variables**

Omits many factors relating to works complexity and treatment levels. No consideration is made of total assets, their age, condition, utilisation, or policies concerning replacement and renewal. *If variables are omitted the residual will contain the effects of omitted factors not correlated with included factors. This corresponds to source C in the list of errors in section 2.1.*

**Poor proxy**

The number of properties is likely to be a very poor proxy. *Poor proxies for explanatory variables lead to errors in variables problems. The coefficients are biased toward zero and the residual contains another term reflecting the error in the explanatory variable, thus reducing the proportion attributable to inefficiency. Corresponds to source D in section 2.1.*

**Sampling error**

The range of adjusted costs and the standard deviation suggest a very large sampling error is present. *Even in the absence of other problems, so that the true disturbance is inefficiency, small samples mean that the equation will not be estimated accurately, leading to errors in the residuals as estimates of relative efficiency. Other sources of error add to this problem. Corresponds to error source A.*

**Measurement error**

There appear to be no particular measurement error problems associated with the dependent variable above those normally associated with the capital maintenance models. *The dependent variable may be measured with error, which a) directly reduces the proportion of the residual attributable to relative efficiency but also b) increases the variance of the model's coefficients. To the extent that the error is one of allocation, there will be some cancelling out across models of effect a), which we will need to consider when discussing the aggregate efficiencies. Corresponds to source B.*

**Mathematical form**

The absence of an equation probably fails to do justice to the complexity of this category of expenditure. *The mathematical form chosen is always a compromise between simplicity, fit and common sense. Some forms may be more susceptible than others to distortions, which may particularly affect companies at the extremes. Corresponds to source E.*

**Evaluation:** One of the poorest models. *An attempt at a one-sentence summary of the model's likely performance.*

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
40	10	35	5	0	10

*This is a key table. The numbers in the boxes should add up to 100%. The individual numbers are derived from the considerations discussed in the boxes above. Efficiency*

*is the residual category. In this example the largest detriment was from omitted variables, whilst mathematical form played no part in reducing the amount of variation attributable to relative efficiency.*

*The individual numbers are derived from the considerations discussed in the boxes above. As this assessment is necessarily qualitative, the variation in the residual is attributed according to the author's opinion based on the available evidence, and analysis of the models.*

Deriving the contribution of each type of error to the total residual is not mechanical, but the following considerations apply:

1. Omitted variables: A higher weight will be given where there is more evidence of omitted variables, depending on their number and likely importance. All the capital maintenance models have a higher weighting because age, condition and utilisation of assets cannot be easily measured and therefore is omitted. Even in a good model which largely reflects industry knowledge 10% or more of the residual could easily be due to omitted factors. The likelihood of the importance of substantive omitted factors is increased where the adjusted cost ratio is particularly high. In a poor model with one major or several smaller excluded factors, 40% or even 50% of the residual could be attributable to omitted factors. The capital maintenance models are particularly affected by this factor.
2. Poor proxy: The weight given to this will increase the more distant the cost driver measure that is used from the ideal economic or engineering concept. Inevitably, observed variables will depart from the ideal, but in a good model this should be in the range 0-5%. However, in some cases the approximation effect could account for as much as 25% of the residual.
3. Sampling error will be greater the smaller the number of observations used in regression or unit cost "boxes" and the greater the standard error of estimate or adjusted cost ratio. For example, the large sewage treatment opex model (369 observations) and sewerage network models (64 observations) have less than 5% of their residual attributed to sampling error. The water models with 22 observations generally have higher sampling errors than the sewerage models with only 10 observations. Despite 22 observations, water resources and treatment capital maintenance has a relatively high sampling error attributed to it (35%) owing to its high ratio of adjusted costs (35:1) and high standard deviation of its residual.
4. Measurement error (in the expenditure variable) will be greater where there is greater scope for variation in the way companies have allocated cost. This allocation uncertainty occurs both within each broad category (operating and capital maintenance) and across activities within the broad categories. Within operating expenditure the scores embody the judgement that business activities, being a residual category, are likely to have the highest variance associated with measurement error. Somewhat similar considerations may apply to the management and general category for the capital maintenance models. The highest score for measurement error is 15%, and this has been applied to these "residual categories" models. More commonly, only 5% or maybe 10% of the residual has been attributed to this category of error.

5. Mathematical form. Larger potential problems with the mathematical form are associated with a larger attribution of error to this component. In most cases this is not more than 5 – 10% of the total disturbance.

Other things equal, the highest adjusted cost ratios carry a presumption that a higher proportion of the observed adjusted cost variation may be due to equation errors of one sort or another, but for each model this needs to be associated with particular reasons.

### 3. Evaluation of each model

We now turn to evaluate Ofwat's models and the weight that they can bear in setting regulated cost allowances.

All the companies we received comments from had reservations about some models. Where companies did well they tended to be less critical, but have tended to focus their attention on models where they did relatively badly. All companies had reservations about the capital maintenance models whether they did well or badly on these models. Ofwat also clearly has some reservations about its own equations. For example, Cambridge Water is sometimes "too small" to be a valid comparator, and it has concerns about Yorkshire Water's methods of cost allocation. By holding discussion groups to consider alternative models Ofwat seems to be making an effort to find the best (and therefore most defensible) model.

The models inevitably represent a compromise amongst the five criteria applied by Professor Stewart – three of them in particular. The first is consistency with engineering knowledge. In this respect the early models based on 1992-3 costs as originally developed by Professor Stewart were arguably the most coherent, albeit limited to operating expenditure. The second is statistical consistency with the pattern of the observed data. Unfortunately several of the early models broke down as time went on.

Companies changed their costs in response to the strong incentives provided by price cap regulation. Some of them may also have different accounting policies in line with their perceived interest given Ofwat's behaviour with respect to capital and operating costs. There may have been a tendency to look particularly hard at areas where they appeared to have relatively high unit costs. They are likely to have also looked more carefully and in some cases changed the way they have measured explanatory factors. These responses may have contributed to the change in the estimated models. Companies are likely to have behaved differently (different cost allocations, and different explanators) making company comparisons more difficult, further leading to model breakdown.

Against this, data which are used in the comparative efficiency models ought to become more accurate over time. Companies will find it worth making more effort to collect the data comprehensively, and auditors will inspect it more carefully. Efforts are made to standardise methods across the industry.

The third element is to use models which do not give companies perverse incentives. For example, for the opex models particularly Ofwat has tended to avoid using assets as explanatory variables lest they give an incentive to keep uneconomic assets in operation. It is obviously less easy to defend this for the capital maintenance models.

Finally the models are limited to using what data can actually be collected on an industry wide basis. This means that the variables used are often just proxies for other variables.

### 3.1 Operating Cost Models: Water

Table 3.1.1 Model Summary

Water Opex	Variation in residual attributable to:					
Water	Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
Distribution	20	25	10	5	10	30
Business Activities	25	5	10	15	2	43
Power	10	5	5	5	5	70
Resources & Treatment	25	20	13	5	7	30

**Activity modelled:** Water distribution operating cost

**Dependent variable:** Ln(Distribution expenditure per head)

**Scale and/or explanatory factors:** Proportion of main > 300mm

**Interpretation of mathematical properties:** The model has the well-known property that companies with a greater total amount of main but the same amount of large main will have higher costs. Furthermore the elasticity of "unit cost" to this ratio is implausibly high, suggesting that the variable is a best a proxy for something else (e.g. urbanisation) and is therefore subject to severe errors in variables problems.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Linear regression	0.222	0.261	25.28	2.73

**Comments from industry knowledge:** The interpretation of the main explanatory variable has changed from a direct costs interpretation to being a proxy for urbanisation. But the variable is poorly correlated with other measures of urbanisation such as population density.

The model provides no allowance for the extra costs of very rural companies. Ofwat remarks that “because the number of rural customers is a small proportion of the total we can safely ignore their influence on costs”. The key issue is whether for some companies the *costs* of serving rural customers are a small proportion of the total. This depends not only on the number of customers, but also the extra cost per customer. The model also makes no allowance for the differing leakage control costs which companies face depending upon their different operating environments

**Sources of error:**

Omitted variables

Area covered, number of connections, ground conditions, bursts, factors affecting leakage costs.

Poor proxy

Proportion of main > 300mm probably a poor proxy for urbanisation

Sampling error

With 22 observations and typical standard errors implies intermediate value for sampling error.

The range of adjusted costs also suggests implausible levels of inefficiency

Measurement error

Opex/capex trade off, allocation with business activities. moderate

Mathematical form: Although the issue is moderately severe in this model, most of effect is probably felt through the poor proxy effect.

**Evaluation:** One of the poorer models in water supply.

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
20	25	10	5	10	30

**Activity modelled:** Water power

**Dependent variable:** Ln (Pumping expenditure)

**Scale and/or explanatory factors:** Ln(Distribution Input \*average pumping head)

**Interpretation of mathematical properties:** A 10% rise in distribution input or pumping head leads to a 9.4% increase in power costs.

No other factors are taken into account.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Linear regression	0.141	0.989	14.38	1.79

**Comments from industry knowledge:** This is one of the better models, and the relatively small standard errors, standard deviation of residuals, and range of adjusted cost reflects this.

**Sources of error:**

**Omitted variables**

Each pumping site incurs standing charges, so number of sites is potentially relevant, as the distribution element in power costs, which varies regionally. Presumably the age, condition, and size of pumps will also affect their efficiency. The model takes no account of power used for other purposes e.g. crypto sporidium monitoring and treatment.

**Poor proxy**

Pumping head figures have been revised over time and should now be a fairly reasonable proxy.

**Sampling error**

Having only 22 observations and a relatively small but still-significant standard deviation of adjusted costs means that sampling error will be moderate.

**Measurement error**

Relatively few allocation issues involved in this variable

**Mathematical form**

Not a major issue

**Evaluation:** One of the best models overall.

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
10	5	5	5	5	70

**Activity modelled:** Water resources and treatment

**Dependent variable:** Resources and treatment spending less power and EA charges divided by population

**Scale and/or explanatory factors:** Number of sources/distribution input , proportion of supply from rivers.

**Interpretation of mathematical properties:** Cost of resources and treatment would be £1.49 if there were no river sources or £6.60 if it all came from rivers plus an additional element inversely proportional to average source size.

The form of the equation implies:

1. A constant cost of £1.49 per head of treatment cost *plus*
2. £5.12 per head if all water comes from rivers (proportionately less otherwise) *plus*
- 3a. 17 pence per person for every extra source if the total distribution input is 100 million litres per day *or*
- 3b. 8 pence per person for every extra source if the distribution input is 200 million litres per day.
- 3c. 4 pence per person for every extra source if the distribution input is 400 million litres per day.

This means that for two companies with the same distribution input of 500Ml/day and 50% river sources, if company A has 100 sources and B has 50 sources, the company A's predicted cost will be £7.40/head and B's will be £5.72/head.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Linear regression	1.859	0.274	31.61	4.08

**Comments from industry knowledge:**

Arguably costs are driven more by the number of treatment plants than the number of sources.

**Sources of error:**

Omitted variables

There is no explicit role for treatment types, not sufficiently captured by river sources.

Poor proxy

1. River sources are a poor proxy for treatment type.
2. Treatment plants rather than sources should perhaps be considered.

There is a danger of double counting omitted variables and poor proxies, and the scores below are modified to take account of this.

**Sampling error**

The SEE is not comparable with other models because of the nature of the dependent variable. The standard deviation of percentage residuals is rather high as is the range of adjusted costs, suggesting that sampling error will be relatively high in this model with 22 observations.

**Measurement error**

The number of sources may be measured differently between companies, as only operation sources are allowed. Companies will have other standby sources which may be treated differently.

**Mathematical form**

The use of population as the deflator in the dependent variable is curious. A population fall of 10% accompanied by a rise in per capita consumption of the same magnitude (and therefore no change in distribution input) would lead to a fall in modelled cost of 10%.

**Evaluation:** One of the poorer models in water supply

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
25	20	13	5	7	30

**Activity modelled:** Water business activities

**Dependent variable:** Ln(Business activities expenditure)

**Scale and/or explanatory factors:** Ln (number of billed properties)

**Interpretation of mathematical properties:** A 10% rise in billed properties leads to a rise of 9.49% in business activities expenditures including doubtful debts but excluding rates.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Linear regression	0.224	0.966	23.22	2.56

**Comments from industry knowledge:** Business activities represent a range of activities not easily modelled with just a few variables. At the margin costs may be allocated into or out of this category.

**Sources of error:****Omitted variables**

Number of samples tested, number of metered customers, determinants of bad debts.

**Poor proxy**

No major problems once omitted variables have been acknowledged.

**Sampling error**

Given the SEE and standard deviation of the residual we should expect this to be of the same order as for the water distribution model.

**Measurement error**

Business activities are more substitutable at the margin form other activities in terms of cost allocation, should we should expect more than average error from this source.

**Mathematical form**

No issues identified.

**Evaluation:** A moderately successful model

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
25	5	10	15	2	43

### 3.2 Operating Cost Models: Sewerage

Table 3.2.1 Summary of models

Sewerage Opex	Variation in residual attributable to:					
Sewerage service	Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
Large Works	20	8	2	7	5	58
Small Works	5	0	25	10	1	59
Network	15	25	5	7	3	45
Business Activities	10	5	40	15	2	28
Sludge treatment and disposal	10	0	40	5	1	44

**Activity modelled:** Sewage treatment at larger works

**Dependent variable:** Ln(Sewage treatment expenditure at large works)

**Scale and/or explanatory factors:** Ln(Total load), tight SS or BOD consent, activated sludge dummies

**Interpretation of mathematical properties:** At each works at 10% increase in load leads to a 7.54% increase in costs. If there is a tight effluent consent for both BOD and suspended solids costs increase by 6%. Activated sludge treatment raises costs by 42% ( $=\exp(0.353) - 1$ ).

No other treatment types are considered.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Linear regression	0.470	0.715	27.68	2.10

**Comments from industry knowledge:** This model captures several essential features of sewage treatment, although it cannot capture every aspect of a complex process.

**Sources of error:****Omitted variables**

The age and condition of the plant will also affect operating costs, as will the degree of utilisation and the composition of the load. Not all treatment types are separately considered. The effect of this may be considerable, as newer works built under quality enhancement drivers may have different operating costs to older works.

**Poor proxy**

Load reflects both the design capacity of the plant and the degree of utilisation, which will have different effects on cost.

**Sampling error**

Although the SEE is comparable with other models, the sampling error is reduced by the availability of large number of observations.

**Measurement error**

There are some issues relating to allocation of costs between treatment plants, but this is fairly minor compared with the smaller plants

**Mathematical form**

The double log formulation is a reasonable approximation at plant level, leading to relatively issues.

**Evaluation:** One of the better sewerage service models.

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
20	8	2	7	5	58

**Activity modelled:** Sewage treatment operating expenditure in smaller works

**Dependent variable:** Annual expenditure/load treated

**Scale and/or explanatory factors:** Five size bands, 10 treatment types

**Interpretation of mathematical properties:** Treatment costs are assumed to depend on treatment type and size only. No cost function is used, simply the industry average for each of 50 cells.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Unit costs in 50 cells	n.a.	n.a.	23.25	1.94

**Comments from industry knowledge:** This model addresses the major cost determinants, though the number of observations in some cells may be too low for satisfactory statistical properties.

(The pattern of cell averages does not always follow a regular form as size or treatment type varies – for example the figure of 10.89 for size band 1 preliminary treatment sea outfalls seems unexpectedly high. In previous years there have been negative values in some cells, and we still have zeros for some cells, although perhaps these are not populated. )

Almost by definition, these will have a minor impact for the industry as a whole, but some companies could be significantly affected.

**Sources of error:**

Omitted variables  
Comparable with the large works models

Poor proxy  
The variables used are reasonable proxies for the true cost drivers.

Sampling error  
The small number of observations per cell is mitigated by the fact that most companies should have observations in several cells, thus reducing the overall sampling error.

Measurement error  
We should expect some allocation problems across smaller works

Mathematical form  
Not an issue with this model

**Evaluation:** A reasonably good model.

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
5	0	25	10	1	59

**Activity modelled:** Sewerage network operating expenditure

**Dependent variable:** Ln(Network Expenditure/km sewer)

**Scale and/or explanatory factors:** Ln (Area/Sewer length), Ln(resident population/sewer length), holiday population/residential population.

**Interpretation of mathematical properties:** This is fairly complex mathematically. Other things equal, costs vary proportionate to sewer length, which is the scale variable. But if the area of the sewerage catchment is higher by 10%, costs are 1.79% higher.

An increase in the total population (resident plus holiday) will increase costs per kilometre by 4.31%.

If the holiday population rise say from 5% to 15% (or 80% to 90%) of the total, expenditure will need to rise by 7%.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Linear regression	0.259	0.456	15.14	1.74

**Comments from industry knowledge:** Sewer length and population seem to be reasonable variables, but the holiday population raises some issues about what it might be a proxy for.

There are a variety of approaches to calculating holiday population: some companies appear to include day visitors, others not.

**Sources of error:**

Omitted variables

There is no allowance for topology, pumping station frequency

**Poor proxy**

Holiday population is probably subject to some error in measurement and may be a proxy for other effects.

**Sampling error**

The residual has a relatively small standard deviation, and the range of relative efficiencies is at the smaller end of the range, and 64 observations will lead to a relatively small sampling error.

**Measurement error**

There are probably some allocation issues for sewerage area costs, mitigated by the larger than average number of observations.

**Mathematical form**

There is potentially a minor effect arising from the nonlinear nature of the function

**Evaluation:** This model is probably around average in its ability to extract information about efficiency .

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
15	25	5	7	3	45

**Activity modelled:** Sludge treatment and disposal operating expenditure

**Dependent variable:** Annual expenditure/amount of sludge

**Scale and/or explanatory factors:** 3 types of farmland, incineration, landfill, composted, land reclamation, other

**Interpretation of mathematical properties:** This is a comparative unit cost model.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Unit costs in 8 cells	n.a.	n.a.	33.34	2.19

**Comments from industry knowledge:** The types of disposal used are appropriate categories for analysing this category of expenditure. The high standard deviation compared with the range between the maximum and minimum suggests that companies are concentrated in the tails of the distribution, and an inspection of the individual adjusted costs confirms this to be the case.

Previous versions of this model implied cost ratios of 5:1 for landfill and 3:1 for conventional land disposal. In PR 99 Ofwat attributed the wide discrepancy in observed costs as mainly due to problems in cost allocation between this activity and sewage treatment. (Anglian and Thames' costs were "much lower than predicted" in sludge disposal and "much higher than predicted in small sewage treatment plants.)

**Sources of error:**

**Omitted variables**

Disposal to land will cost more where there is a shortage of suitable land for disposal. The dispersal of the works might be expected to lead to different collection costs. The age and condition of incineration works might be expected to bear on incineration operating expenditures. There are suspicions that the effect of this may be large, as newer assets built under quality enhancement drivers may have different operating costs to older works.

**Poor proxy**

The variables used are reasonable proxies.

**Sampling error**

Clustering of the companies in the tails suggests that sampling error would be important in this model. an average of 10 observations per cell will also increase sampling error, offset by the fact that each company will be represented by several observations.

**Measurement error**

Allocation difficulties between this activity and sewage treatment suggests that errors in the dependent variable are potentially significant.

Mathematical form  
No major issues of mathematical form have been identified.

**Evaluation:** A moderately useful model.

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
10	0	40	5	1	44

**Activity modelled:** Sewerage business activities

**Dependent variable:** Annual average expenditure per property

**Scale and/or explanatory factors:** Weighted industry average

**Interpretation of mathematical properties:** This is a unit cost model, implying a constant cost for every property billed.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Simple unit cost	n.a.	n.a.	35.59	2.68

**Comments from industry knowledge:** Although different types of property might have different costs, the effect on the model of excluding these is likely to be moderate.

**Sources of error:**

Omitted variables

Although different types of property might have different costs, the effect on the model of excluding these is likely to be moderate.

Poor proxy

Based on consistency with industry knowledge including accuracy of data for variable used, qualified by goodness of fit, and stability of model over time

Sampling error

With only 10 observations sampling error is likely to have a noticeable impact on the accuracy of the results. This is supported by the relatively high standard deviation for the % residual and the above average value for the adjusted cost ratio.

**Measurement error**

There are some allocation issues at the margin for this model which will increase measurement error.

**Mathematical form**

No major issues arise concerning the mathematical form, although there may be case for allowing for some economies of scale in this activity owing to the presence of fixed functions such as regulation.

**Evaluation:** Despite the absence of major omissions or distortions, sampling variation and measurement error will affect the likely performance of this model.

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
10	5	40	15	2	28

## CAPITAL MAINTENANCE MODELS

**3.3 Capital Maintenance Expenditure Models: Water**

Table 3.3.1 Summary of models

Water Capital Maintenance	Variation in residual attributable to:					
Water	Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
Water Distribution Infrastructure	40	15	5	15	5	20
Water Distribution Non-infrastructure	35	10	15	10	2	28
Water Resources & Treatment	35	10	35	5	0	15
Water Management & General	30	15	25	15	1	14

**Activity modelled:** Water distribution infrastructure capital maintenance

**Dependent variable:** Ln(distribution infrastructure expenditure/length of main)

**Scale and/or explanatory factors:** Ln(Connected properties/total main length)

**Interpretation of mathematical properties:** A 10% rise in the length of main and properties connected will lead to a 10% rise in required expenditure.

Keeping sewer length the same but doubling the number of properties connected raises requires expenditure by 85%.

Raising main length by 10% but keeping the number of properties the same raises required expenditure by around 1%.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Linear regression	0.569	0.496	19.66	2.34

**Comments from industry knowledge:** The number of properties is the key driver in this model rather than length of main. The burst rate, the proportion of communication pipes made of lead (nor proportion of small pipes) no longer appear in this model.

The model has reverted to the 1999 model except that the denominator is now mains length rather than distribution input less estimated leakage.

**Sources of error:**

**Omitted variables**

There is little direct influence of the amount of infrastructure assets, or the need for expenditure, as reflected in the age, condition, or composition of the assets.

**Poor proxy**

The small role played by mains length compared with the number of properties connected suggests that the key variable may be a poor proxy.

**Sampling error**

The sampling error should be moderate, given the slightly below average value for the standard deviation of the % residual. (SEE is not relevant given the nature of the dependent variable.)

**Measurement error**

As with all the capital expenditure models actual expenditure is likely to diverge considerably from required expenditure because of varying company policies about long term asset maintenance and replacement, but is worse for infrastructure assets which are typically longer lived than non-infrastructure assets.

**Mathematical form**

Potentially of some significance, but difficult to establish given the implausibility of the model overall

**Evaluation:** Likely to be a poor performer overall, even for a capital maintenance model.

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
40	15	5	15	5	20

**Activity modelled:** Water distribution non-infrastructure

**Dependent variable:** Ln(non-infrastructure expenditure/pumping station capacity)

**Scale and/or explanatory factors:** Ln(service reservoir and water storage capacity/pumping station capacity)

**Interpretation of mathematical properties:** A doubling of pumping and service storage capacity together leads to a doubling of required expenditure.

A 10% rise in pumping capacity alone leads to a 3% rise in required expenditure.

A 10% rise in local storage leads to a 6.5% rise in required expenditure.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost	Range Excluding Portsmouth
Linear regression	0.569	0.338	63.06	7.78	6.13

**Comments from industry knowledge:** This model implies that storage and pumping capacity are relevant in the ratio of approximately 2:1 for non-infrastructure expenditure. The effect of meters is not included.

**Sources of error:**

**Omitted variables**

The age, condition, and degree of utilisation of assets are not part of the model. The size of the site at which an asset is located (or proportion of smaller sites) may be a relevant cost driver, as may company policy about asset repair versus renewal. Omitted variables are therefore fairly important in this model.

**Poor proxy**

The included variables constitute the major assets of relevance and so this effect is not likely to be of major significance in this model.

**Sampling error**

The sampling error and range of adjusted costs is exceptionally large in this model, so sampling error is quite important.

**Measurement error**

One company was particularly concerned about cost allocation between mains (infrastructure) and pumping stations (non-infrastructure)

**Mathematical form**

No major issues identified

**Evaluation:** A mediocre model, reflected in completely implausible implied efficiency range, even after adjustment.

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
35	10	15	10	2	28

**Activity modelled:** Water Resources and Treatment

**Dependent variable:** Expenditure per property

**Scale and/or explanatory factors:** Weighted industry average

**Interpretation of mathematical properties:** Required expenditure on resources and treatment capital maintenance is the same per connected property for every company.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost	Range excluding Cambridge
Simple unit cost	n.a.	n.a.	48.79	35.02	8.11

**Comments from industry knowledge:** This model is too simple to measure the costs for a complex and diverse set of processes. There are probably insufficient observations to arrive at a satisfactory model. The number of properties is only indirectly associated with higher resource and treatment costs.

**Sources of error:**

**Omitted variables**

Omits many factors relating to works complexity and treatment levels. No consideration is made of total assets, their age, condition, utilisation, or policies concerning replacement and renewal.

**Poor proxy**

The number of properties is likely to be a very poor proxy.

**Sampling error**

The range of adjusted costs and the standard deviation suggest a very large sampling error is present.

**Measurement error**

There appear to be no particular measurement error problems associated with the dependent variable above those normally associated with the capital maintenance models.

**Mathematical form**

The absence of an equation probably fails to do justice to the complexity of this category of expenditure.

**Evaluation:** One of the poorest models.

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
35	10	35	5	0	15

**Activity modelled:** Water Management and General

**Dependent variable:** Ln(Expenditure per property)

**Scale and/or explanatory factors:** Proportion of properties that are non-household

**Interpretation of mathematical properties:** If the number of billed properties, household and non-household, doubles, then required expenditure will double.

The cost per property depends in the following way on the proportion of properties that are non-household:

Proportion Non -Households	Cost/property £
0.00	3.0
0.05	5.3
0.10	9.2
0.15	15.9
0.20	27.7
0.50	762.2

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Linear regression	0.370	0.203	35.23	4.92

**Comments from industry knowledge:** The number of connected properties, household, or other has only an indirect bearing on requirements for capital maintenance in this category.

The impact of non-households seems on the high side.

No allowance is included for possible economies of scale in headquarters activity.

**Sources of error:**

**Omitted variables**

We might expect a large number of small influences here.

**Poor proxy**

The variables used here seem to be poor proxies, given the indirect relationship of the explanatory variable with the activity.

**Sampling error**

There is a large range of observed values and standard deviation of the residual, leading us to expect a high sampling error.

**Measurement error**

Management and general expenditure is more subject than most to variation from allocation decisions.

**Mathematical form**

This is potentially a problem depending on the range of non-household proportions observed.

**Evaluation:** A relatively poor model.

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
30	15	25	15	1	14

### 3.4 Capital Maintenance Expenditure Models: Sewerage

Table 3.4.1 Model summary

Sewerage Capital Maintenance	Variation in residual attributable to:					
Sewerage service	Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
Sewerage Infrastructure	30	20	2	5	10	33
Sewerage Non-infrastructure	20	15	40	5	0	20
Sewage Treatment	45	10	4	5	2	34
Sludge Treatment & Disposal	35	0	35	5	0	25
Sewerage Management & General	25	5	30	15	2	23

**Activity modelled:** Sewerage Infrastructure Capital Maintenance

**Dependent variable:** Ln(expenditure/ length of sewer)

**Scale and/or explanatory factors:** Combined sewer overflows/km sewer, proportion of critical sewers

**Interpretation of mathematical properties:** Doubling sewer length and the number of combined sewer overflows leads to a doubling of cost. An increase in the proportion of sewer that is critical by five percentage points raises required expenditure by 9.5%. A doubling in the number of combined sewer overflows with a given length of sewer raises total costs by 29%.

In the limit, a company which had no combined sewer overflows would have no required expenditure.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Linear regression	0.457	0.427	24.81	2.02

**Comments from industry knowledge:** Both foul and combined sewers contribute to capital maintenance expenditure requirements, but the representation in this equation is a little strange.

**Sources of error:**

**Omitted variables**

No account taken of policies relating to asset replacement, refurbishment and renewal. One company claims there is "a strong correlation between the percentage of MEAV replaced and the model residual." Soil conditions and urban environments may be additional influences.

**Poor proxy**

The volume and type of assets are appropriate variables. However, there is some uncertainty over the comparability of the "critical" sewer criterion across companies. CSOs are only an indirect measure of the length of combined sewers.

**Sampling error**

The higher than average standard deviation of % residual is offset by the larger number of observations than in most other models though from only 10 companies.

**Measurement error**

There are likely to be significant allocation uncertainties between different sewerage areas. Earlier versions of this equation omitted Severn-Trent sewerage areas owing to differences in definitions.

**Mathematical form**

If there are any companies with very small proportions of CSOs the distortion could be significant

**Evaluation:** The attempt to take account of asset condition improves the model. Although it is not one of the worst capital maintenance models and has the lowest range of adjusted costs, it is still marred by a number of shortcomings.

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
30	20	2	5	10	33

**Activity modelled:** Sewerage non-infrastructure capital maintenance expenditure

**Dependent variable:** Expenditure/pumping station

**Scale and/or explanatory factors:** Weighted industry average

**Interpretation of mathematical properties:** Expenditure requirements are proportional to the number of pumping stations.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Simple unit cost	n.a.	n.a.	48.79	3.53

**Comments from industry knowledge:** This model is too simple to capture the factors affecting non-infrastructure requirements.

**Sources of error:****Omitted variables**

Ignores size, type, age, utilisation, condition of assets, or company policies concerning renewal or refurbishment of assets.

**Poor proxy**

The number of pumping stations is only an approximation to the volume of assets to be maintained.

**Sampling error**

A high standard deviation of the % residual and adjusted cost ratio with only 10 observations suggests a substantial sampling error. One company (Anglian) significantly affects the average cost. See chart A3.1 in appendix A3.

**Measurement error**

Typical values for a capital maintenance model.

**Mathematical form**

There are no issues arising from the mathematical form of this model.

**Evaluation:** This model fails to capture many potential influences on expenditure requirements, and is below average for capital maintenance models.

Variation in residual attributable to:

Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
20	15	40	5	0	20

**Activity modelled:** Sewage Treatment Capital Maintenance Expenditure

**Dependent variable:** Ln(expenditure/ total load received)

**Scale and/or explanatory factors:** Ln(number of works/total load)

**Interpretation of mathematical properties:** Doubling the number of works and total load doubles modelled expenditure requirements. Reducing load per works by 50% raises unit expenditure requirements by 12%

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Linear regression	0.522	0.210	51.21	3.75

**Comments from industry knowledge:** This model simply reflects average works size. Treatment types are no longer included in the model.

**Sources of error:**

**Omitted variables**

Makes no references to treatment type and complexity, or asset age, condition or utilisation. These omissions are likely to have a substantial impact on the efficiency component in the residual. Newer works built under quality enhancement drivers may have lower maintenance requirements than older works.

**Poor proxy**

Based on consistency with industry knowledge including accuracy of data for variable used, qualified by goodness of fit, and stability of model over time

**Sampling error**

The large standard deviation and adjusted cost range suggests that sampling error would be significant, mitigated by the relatively large number of observations.

**Measurement error**

Average for capital maintenance models

**Mathematical form**

Only minor issues would be expected arise in this model

**Evaluation:** A model of average quality amongst the capital maintenance models.

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
45	10	4	5	2	34

**Activity modelled:** Sludge Treatment & Disposal

**Dependent variable:** Expenditure/weight of dry solids

**Scale and/or explanatory factors:** Weighted industry average

**Interpretation of mathematical properties:** Required expenditure is proportional to the total weight of dry solids.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Simple unit cost	n.a.	n.a.	48.73	12.43

**Comments from industry knowledge:** This model is too simple to reflect the determinants of required capital maintenance expenditure for this activity

**Sources of error:**

**Omitted variables**

Route of disposal is ignored, as are the age and condition of assets involved. The age of assets may vary considerably owing to the recent needs of some companies to find alternatives to sea disposal. Newer assets built under quality enhancement drivers may have lower maintenance requirements than older assets.

**Poor proxy**

The scale variable used is appropriate but only reflects one aspect of the processes involved.

**Sampling error**

The very high standard deviation of % residuals and the exceptionally high adjusted cost ratio combined with only 10 observations suggests that sampling error would be an important issue for this model.

**Measurement error**

The measurement is subject to the usual errors in capital maintenance models.

**Mathematical form**

There are no issues concerning the mathematical form.

**Evaluation:** Omitted factors and sampling error make this a worse than average model.

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
35	0	35	5	0	25

**Activity modelled:** Sewerage Management and General Capital Maintenance Expenditure

**Dependent variable:** Expenditure/billed property

**Scale and/or explanatory factors:** Weighted industry average

**Interpretation of mathematical properties:** Expenditure requirements are proportional to the number of properties billed.

**Descriptive statistics:**

Mathematical form	SEE	R <sup>2</sup>	S.D of % residual	Range of adjusted cost
Simple unit cost	n.a.	n.a.	40.66	3.69

**Comments from industry knowledge:** This model is extremely simple, reflecting the fewness of observations relative to the influences that may be affecting individual company spend.

For example, no distinction is made between household and non-household properties billed, in contrast to the water services model for this activity. No allowance is made for possible economies of scale in headquarters activity. (10 observations will not usually provide very strong tests for constant returns to scale, especially with large residual errors. This will tend to lead to a high probability of type 2 error - falsely accepting the null hypothesis)

**Sources of error:****Omitted variables**

The type of property being billed, age and condition of assets.

**Poor proxy**

The number of properties billed only reflects one aspect of scale for a sewerage company, since water only companies in the area typically bill customers themselves for sewerage services.

**Sampling error**

The large standard deviation of the percentage residual indicates significant sampling error with only 10 observations

**Measurement error**

As a residual category, management and general is particularly susceptible to measurement error.

**Mathematical form**

No particular issues arise, apart from the possibility that the imposition of constant return distorts the results.

**Evaluation:** Simplicity brings errors from a number of sources, implying a low information content on relative efficiency.

Variation in residual attributable to:					
Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency
25	5	30	15	2	23

## 4. Aggregation issues

Having derived the contributions to efficiency measurement for each of the individual models, we can now assess how much the models contribute to efficiency assessment overall for each of the four main categories of expenditure.

### 4.1 Problems of Aggregation

Aggregation is a complex issue and potentially fraught with error. For example, the properties of the aggregate will depend, amongst other things, on the degree of correlation of the components of the residual across activities. Even where there is low or zero correlation there will tend to be some cancelling out of errors in the averaging process. This effect can be seen in the fact that the inefficiency estimates for the aggregate expenditure categories are much smaller than the individual estimates.

A natural assumption would be that the degree of correlation is the same for all the components of the residual. This would imply, by a symmetry argument, that the proportion of the residual attributable to inefficiency is unchanged by the aggregation process.

However, this assumption needs some adjustment to take account of the fact that, for one of the categories of error, there is a negative correlation across models. This is

the case for that part of measurement error which is due to arbitrary allocation at the margin.

Where there are differences in allocation policies across companies within a category of spending this will lead to

- a) on average an overestimate of a company's efficiency in activity 1 being offset by on average an underestimate in activity 2.
- b) a greater degree of imprecision in the model coefficients leading to greater sampling error.

While we can do nothing about b) we can make allowance for a). We do this below by assuming that 60% of the effect of measurement error is cancelled out in the aggregation process. This may well be an overstatement of the cancelling out effect, as it is not based on detailed study.

Purely as a sensitivity analysis, we also considered the effect of making the extreme assumption that 90% of the measurement error cancels out.

We have not made any adjustment for either trade off or differing cost allocation between operating and capital maintenance expenditure, which is potentially a major problem.

The results are shown in Tables 4.1 - 4.4 below.

**Table 4.1 Water Opex Aggregation**

	Variation in residual attributable to:					Efficiency			Weight £m
	Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency	adding back 60% of measurement error	adding back 90% of measurement error	
Water Opex									
Distribution	20	25	10	5	10	30	33	34.5	415
Business Activities	25	5	10	15	2	43	47	51.5	325
Power	10	5	5	5	5	70	73	74.5	114
Resources & Treatment	25	20	13	5	7	30	28	29.5	303
					<b>Wtd Av</b>	<b>38</b>	<b>42</b>	<b>45</b>	

**Table 4.2 Sewerage Opex Aggregation**

Sewerage service Opex	Variation in residual attributable to:					Efficiency			Weight £m
	Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency	adding back 60% of measurement error	adding back 90% of measurement error	
Large Works	20	8	2	7	5	58	67.2	69.3	183
Small Works	5	0	25	10	1	59	66	69	125
Network	15	25	5	7	3	45	49.2	51.3	185
Business Activities	10	5	40	15	2	28	37	41.5	256
Sludge treatment and disposal	10	0	40	5	1	44	53	54.5	264
					<b>Wtd Av</b>	<b>45</b>	<b>50</b>	<b>53</b>	

**Table 4.3 Water Capital Maintenance Aggregation**

Water	Variation in residual attributable to:					Efficiency			Weight
	Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency	adding back 60% of measurement error	adding back 90% of measurement error	
<b>Capital maintenance</b>									£m
Water Distribution Infrastructure	40	15	5	15	5	20	24	28.5	234
Water Distribution Non-infrastructure	35	10	15	10	2	28	29	32	126
Water Resources & Treatment	35	10	35	5	0	15	13	14.5	195
Water Management & General	30	15	25	15	1	14	23	27.5	184
					<b>Wtd Av</b>	<b>19</b>	<b>25</b>	<b>29</b>	

**Table 4.4 Sewerage Capital Maintenance Aggregation**

Sewerage service Capital maintenance	Variation in residual attributable to:					Efficiency			Weight £m
	Omitted variables	Poor proxy	Sampling error	Measurement error	Mathematical form	Efficiency	adding back 60% of measurement error	adding back 90% of measurement error	
Sewerage Infrastructure	30	20	2	5	10	33	36	37.5	175
Sewerage Non-infrastructure	20	15	40	5	0	20	23	24.5	51
Sewage Treatment	45	10	4	5	2	34	32	33.5	278
Sludge Treatment & Disposal	35	0	35	5	0	25	23	24.5	62
Sewerage Management & General	25	5	30	15	2	23	32	36.5	161
					<b>Wtd Av</b>	<b>30</b>	<b>34</b>	<b>36</b>	

It might also be argued that we should expect the efficiency component of the models to be positively correlated across the categories. The effect of this would be to raise the share of relative efficiency in the aggregate model.

Against this, there is an argument that the omitted factors will also be positively correlated across expenditure categories, perhaps to the same degree as the error component. This would tend to cancel out the effect of any common efficiency component.

## 4.2 Summary of findings

Table 4.5 brings together the results from Tables 4.1- 4.4. It can be seen that the contribution of efficiency to the estimated residuals varies considerably across the sectors. The operating cost models are better than the capital maintenance models.

Slightly less expected, however, is the fact that the sewerage models come out better than the water models in aggregate.

**Table 4.5 Summary of findings**

	Estimated proportion of residual attributable to efficiency (point estimates)
<b>Operating cost</b>	
<i>Water</i>	42
<i>Sewerage</i>	50
<b>Capital maintenance</b>	
<i>Water</i>	25
<i>Sewerage</i>	34

Note: These should be rounded at least to the nearest 5%, as in the introductory sections of the paper. They are given in unrounded form here as a matter of record.

Some of this is due to the application of weighting factors according to the total share of expenditure. For example the best model is the water power model, but as this has a relatively low expenditure it contributes less than a quarter to the overall water opex score.

## 5. Limitations and potential questions

This paper has used a particular methodology to try to assess in a systematic and reasoned way the contribution of relative efficiency to the residuals in Ofwat's efficiency models.

The resulting estimates are the results of a large number of judgements each of which is subject to error and we would not claim at this stage that these numbers are the final word on model accuracy. Although we believe that they are generally of the correct order of magnitude we are happy to consider further evidence which may enable a more accurate assessment to be made.

In this section we consider some possible questions and objections (shown in **bold**.)

**1. Why are the results much lower than Ofwat's estimate of the proportion of variation due to inefficiency?**

Our results for *sampling* error are not very different from Ofwat's estimate of the total error - of the order of 10% for water opex and 20% for the other categories. But our estimates take explicit account of the other types of error. Since Ofwat has not provided a description of its method of arriving at its figure it is possible that it has ignored the other sources of error.

**2. But arguably the "omitted variables" error should not be included. Ofwat tests for omitted variables, allows for special factors, and, by excluding it "stops a very small company with an especially benign working environment introducing errors into this aspect of our work."**

The tests for omitted variables include only those that can be adequately measured. Ofwat allows that the number of variables it can include is limited because of the number of observations. "There are other explanatory variables which may explain differences in costs but we have found that these are either impossible to measure or difficult to do so on a consistent basis."

With only 10 or 20 observations for many models the power of the hypothesis test used is not very high. (i.e. the probability is relatively high of drawing a sample where a particular variable is not statistically significant although it has a relevant impact.)

Ofwat's acceptance of special factors and the exclusion of the one company with the "especially benign working environment" is tacit acceptance that despite the findings of significance tests, omitted variables are a potential problem.

As an example, we appear to be in agreement with Ofwat that the age and condition of assets is a common problem for all the capital maintenance models: "There are no objective and consistent measures of asset inheritance. This means that it is not possible to take this into account in the econometric analysis."

This suggests that they should take account of it in their allowance for model errors.

**3. How do your figures compare with other estimates?**

At present Ofwat claims an accuracy of 90% for its water models and 80% for its sewerage models. As far as I am aware Ofwat has not described its method of deciding on these figures. (For example, their work on stochastic frontiers would provide an objective estimate<sup>7</sup>, admittedly imperfect.)

The relative ranking of the water and sewerage results suggests that it may be based on the sampling component of the error. Furthermore, the absence of a distinction between operating and capital maintenance expenditure suggests that consideration of

---

<sup>7</sup> Stochastic frontier estimates involve a parameter which is the ratio of one-sided to total error. This "gamma ratio" is often taken to be a measure of the relative importance of efficiency effects versus other sources of equation error.

the important omitted variable of age and condition of assets did not play a role in determining the discount applied to the residual.

#### 4. Do your estimates allow for “swings and roundabouts” arguments?

This is the argument that errors in one model will cancel out with errors in other models.

There are two places where there are potential swings and roundabouts. Firstly, with cost allocation within categories of operating costs. For this we have allowed a substantial part of measurement error to be cancelled out across the models in aggregating the results for the individual models, reflecting the averaging process used by Ofwat in combining models within each major expenditure heading.

The second case refers to the allocation between operating costs and capital maintenance expenditures. In fact here there is no cancelling out; on the contrary, companies are subject to “double jeopardy”.

**Table 5.1**

Company	Q	Capital maintenance	Operating expenditure	Cost	Appropriate target	Double jeopardy target
A	50	70	<b>36</b>	106	100	70.71
B	50	60	42	102	100	70.71
C	50	50	50	100	100	70.71
D	50	40	63	103	100	70.71
E	50	<b>35</b>	71	106	100	70.71

Table 5.1 shows five companies which have made different judgements about how to minimise costs. All are assumed to be technically efficient but only C is allocatively efficient. Their production function is very simple: there are no explanatory factors and output is given by the standard Cobb-Douglas production function

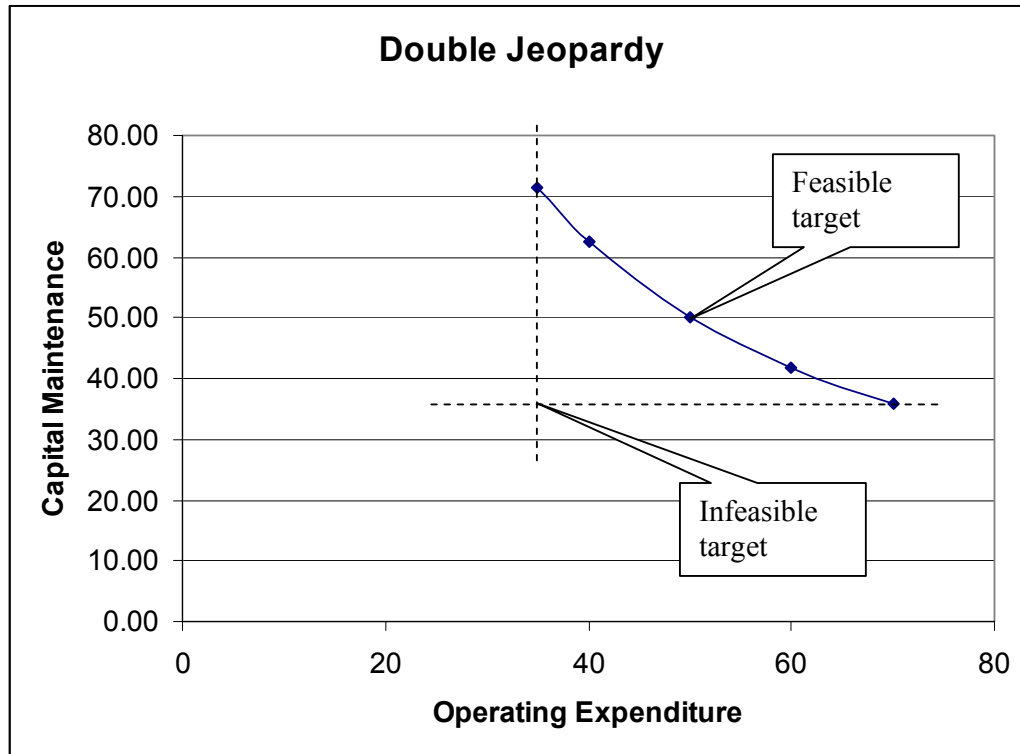
$$Q = Opex^{0.5} Capex^{0.5}$$

Where  $Q$  = total quantity. The unit price of opex and capital maintenance is assumed to be 1.00.

What should the regulatory authority do? The least cost company is C and all companies should be given a target of reaching costs of 100 over a period during which they can adjust their factor ratios (**assuming perfect efficiency models!**).

Now suppose the regulatory authority targets operating and capital maintenance expenditure separately. All companies' targets are 36 for operating expenditure (Company A is the “most efficient”) and 35 for capital maintenance expenditure (Company E is the “most efficient”). But we know by construction that this is not technically feasible! The frontier is the solid line in figure 5.1 (the “isoquant”) and *not* the pair of broken lines, which corresponds roughly to Ofwat's general practice.

Figure 5.1



Ofwat has been aware of this issue for at least ten years. Its current position seems to be:

“We recognise this is a serious issue” and

“At the moment our research tell us that business decisions are not contingent on the interplay with regulatory incentives and that they stem from the same thought and analysis processes that are extant in the competitive market.”<sup>8</sup>

They appear not to have put forward any proposals for dealing with this issue.

##### 5. Companies have outperformed their targets in the past. Does this not validate the regression approach?

Aggregate out-performance is a feature of price cap regulation which has little to do with *comparative* efficiency. The weak correlations observed between efficiency classification and subsequent cost reductions are equally, if not more, consistent with the levels of error in the models discussed here than the 10% and 20% figures put forward by Ofwat.

<sup>8</sup> *Setting Water and Sewerage Price Limits for 2005-10: Framework and Approach* 27 March 2003.

## 6. What is the range of efficiencies implied by the assessment? Are they plausible?

Table 5.2 shows the implications of scaling down the adjusted cost ratio by the percentage of the residual attributed to efficiency. For example if Ofwat's original analysis suggests that the highest-cost company is 90% dearer than the least-cost company (i.e. a range of 1:90) and our analysis indicates that 40% of the residual is inefficiency, then our calculation shows an excess cost of 90% times 40% = 36%. The adjusted cost range in Table 5.2 would then be 1.36, suggesting that the least efficient company could reduce its costs by 36 in 136 or by 26%. (This assumes that the proportion of the residual which is due to efficiency is the same for every company, and that the least cost company sets the frontier.)<sup>9</sup>

**Table 5.2 Adjusted cost range implied by estimates**

Operating expenditure	Ofwat's original range	% attributed to efficiency	Implied adjusted cost range
<b>Water</b>			
Distribution	2.73	30	1.52
Business Activities	2.56	43	1.67
Power	1.79	70	1.55
Resources & Treatment	4.08	30	1.92
<b>Sewerage service</b>			
Large Works	2.10	58	1.64
Small Works	1.94	59	1.55
Network	1.74	45	1.33
Business Activities	2.68	28	1.47
Sludge treatment and disposal	2.19	44	1.52
<b>Capital maintenance</b>			
<b>Water</b>			
Water Distribution Infrastructure	2.34	20	1.27
Water Distribution Non-infrastructure	7.78	28	2.90
Water Resources & Treatment	35.02	15	6.10
Water Management & General	4.92	14	1.55
<b>Sewerage service</b>			
Sewerage Infrastructure	2.02	33	1.34
Sewerage Non-infrastructure	3.53	20	1.51
Sewage Treatment	3.75	34	1.94
Sludge Treatment & Disposal	12.43	25	3.86
Sewerage Management & General	3.69	23	1.62

<sup>9</sup> The formula used is; adjusted cost ratio =  $1 + (\text{Ofwat's cost ratio} - 1) \times \alpha$ , where  $\alpha$  = proportion of residual attributed to inefficiency.

These range from a minimum of 1.33 to a maximum of 6.10 for the individual activities with an unweighted average of 2.0. (This would imply, for instance that the least efficient company had a unit cost 100% higher than the least-cost company.)

The lowest figure, for sewerage network operating expenditure, implies that the least efficient company has unit costs 33% higher than the least cost company. Given Ofwat's belief that "there may be *up to* a 40% difference between the most and least efficient companies" this range might be considered consistent with Ofwat's stated view for each company as a whole. Conversely, a cost ratio of 6.1 might be considered to be implausibly high.

Without making a judgement about company efficiency on the individual activities it is not possible to be precise about what figure would be correct, but most of the resultant ratios might be considered to be a move in the direction of the "less implausible" range compared with Ofwat's own estimates for the individual activities.

## **7. How should companies and Ofwat use these assessments?**

There may be many ways of applying the results. The following is based upon making the smallest changes to Ofwat's methodology consistent with reflecting the issues raised here.

The assessments need to be considered in the context of the whole price control review. In general one would wish to consider possible asymmetries in the consequences of over- and under- estimating the scope for efficiency savings, which is beyond the scope of the present exercise. As Section 6 below shows, sampling and other equation errors lead to misclassification of companies, which means that the results need to be applied cautiously. On average the efficiency range is exaggerated, but the ranking of companies (i.e. relative to each other) will be affected as well.

Suppose one takes as a first order approximation that one is indifferent to under- and over-estimating efficiency potential, and that the scope for improving the models with the data likely to be available is very limited.. Then the appropriate approach to the assessment estimate of efficiency is to scale down the residuals in the existing models much as we have done in Table 5.2 for the adjusted cost ratios. This is essentially the same way that Ofwat has done it, except with more realistic figures for the discount factor - i.e. the proportion of the residual which can be attributed to factors other than variations in efficiency. (There may be a question of whether it is better to scale down the individual model residuals or the aggregate residual, and this would which would require some consideration as there are arguments for doing each. If one were to do the former, some adjustment for the swings and roundabouts effects might be needed.)

Secondly one would need to add together the (adjusted) residuals for the operating expenditure and capital maintenance models in the same way that Ofwat adds the result of the individual models for the separate aggregates. Again, as Ofwat does for each of the four expenditure groups, after eliminating outliers one could then deduce an overall benchmark company for water and one for sewerage. The major difference

from what we have at present would be that we would not have separate aggregates for operating and capital maintenance activities. This would deal with the problem of double jeopardy at a stroke without the need to estimate a frontier representing the trade off between operating and capital maintenance expenditure.

Finally, in order to enter the results into the financial model, one would need to consider the period over which it is reasonable to assume that efficiencies could be realised, and this is beyond the scope of this project.

This is just an indication of one way in which the results could be used in a mechanical way as a one-shot device. One would not wish to create expectations that this method would become the standard approach in future price reviews, as it would create incentives to distort cost allocations away from activities where the models are better to where they are poorer.

## **6. The consequences of error: a simulation**

In applying the above results, there is a balance to be struck between making Type 1 and Type 2 errors – falsely classifying a company as efficient and falsely classifying it as inefficient. Ideally, rather than adopting a simple mechanical rule as described in Section 5, this requires a careful consideration of the costs and probabilities associated with each of these errors, which is beyond the scope of the present paper.

As a contribution to the analysis, however, we have calculated the probability of misclassification for different scenarios for the case of 22 observations.

This involved simulating a data generation process (“Monte Carlo analysis”) using Excel, which is reported in Table 6.1. The cost driver was assumed to be a standard normal deviate, there was a constant term of 10 and a slope parameter of 1.00. All errors were assumed to be independently normally distributed with zero mean. Different scenarios were carried out by changing the standard deviations of each of the error components.

All the results are based on 500 replications. This is not large by Monte Carlo standards, but was sufficient to obtain convergence to three significant figures. Columns 2-4 show how the error variance is composed for the replications reported, and column 5 shows the proportion of the overall variance due to inefficiency.

Columns 7, 8 and 9 report on the standard regression parameters. Columns 10 and 11 report on the adjusted cost ratios as we have calculated them for Ofwat's models – column 10 is the unadjusted ratio and column 11 is the adjusted ratio as we have calculated it in Section 5 above. (Our simulations have a slightly lower range than Ofwat's).

Column 12 shows the proportion of companies which are misclassified into efficiency bands A-E. Column 13 shows the proportion which are misclassified by more than 1 band.

Row 1 shows a base case where all the disturbance is due to variations in efficiency, but the disturbance is very small. The standard error of the estimate is very small at 1.5%, the t ratio is over 400 and there is no misclassification reported in any of the estimated equations.

Row 2 shows what happens when the disturbance is still pure efficiency, but the variance is larger. For the remaining rows the standard deviation of the total error was set at 20% i.e. an overall error variance of 0.04. The raw adjusted cost ratio is just over 2 and for Row 2 the standard error of estimate is around 18%. The result is some misclassification (31%) and 2% of companies are “seriously” misclassified by two bands.

As the proportion of the error which is due to factors other than inefficiency rises the degree of misclassification also rises. Rows 3 and 4 show Ofwat's preferred values of 90% (for opex) and 80% (for capex). These give rise to misclassification rates of 41% and 47% respectively.

Rows 5-8 shows what happens for the values we have estimated for the proportion of variation attributable to efficiency. It can be readily seen that misclassification rises steadily with the reduction in the accuracy of the residual as an estimate of efficiency.

A simple reading of these results, taken together with our estimates of the proportion of the disturbance term which is due to efficiency is that, in applying the results of its efficiency analysis, Ofwat needs to take account of the high likelihood of misclassification of companies into efficiency bands and the consequential impact this has on efficiency targets incorporated in price limits. This applies even at the 80% and 90% figures for efficiency content in the residual which Ofwat has proposed.

**Table 6.1 Simulations of different disturbance components**

Parameters					Results							
(1) Const	(2) s.d. of errors in variables	(3) s.d. of omitted variables	(4) s.d. of Measurement error in y	(5) s.d. of efficiency	(6) Efficiency	(7) R Sqd	(8) t b hat	(9) Std error estimate	(10) Residual range (adjusted cost ratio)	(11) Efficiency range (adjusted cost ratio)	(12) Misclassification by 1 or more classes	(13) Misclassification by 2 or more classes
1	0.00	0.00	0.00	0.01	<b>100%</b>	0.9999	468.497	0.015	1.04	1.038	<b>0%</b>	0%
2	0.00	0.00	0.00	0.20	<b>100%</b>	0.961	23.9	0.175	2.13	2.122	<b>31%</b>	2%
3	0.03	0.04	0.03	0.17	<b>90%</b>	0.9692	26.963	0.152	1.93	1.807	<b>41%</b>	4%
4	0.04	0.04	0.05	0.15	<b>80%</b>	0.9718	28.285	0.188	1.88	1.665	<b>47%</b>	6%
5	0.00	0.12	0.07	0.14	<b>50%</b>	0.9592	23.507	0.209	2.13	1.482	<b>62%</b>	20%
6	0.00	0.14	0.07	0.13	<b>40%</b>	0.9598	23.553	0.205	2.15	1.376	<b>66%</b>	22%
7	0.00	0.15	0.07	0.12	<b>35%</b>	0.9591	23.188	0.249	2.15	1.336	<b>68%</b>	23%
8	0.00	0.16	0.07	0.10	<b>25%</b>	0.9593	23.368	0.197	2.13	1.237	<b>73%</b>	27%

Note: Five efficiency groups were defined with boundaries at -15%, -5%, +5%, and +15%  
 Since there are five efficiency classes, a purely random allocation would be expected to have a misclassification rate of around 80%.

A simple reading of these results, taken together with our estimates of the proportion of the disturbance term which is due to efficiency is that, in applying the results of its efficiency analysis, Ofwat needs to take account of the high likelihood of misclassification of companies into efficiency bands and the consequential impact this has on efficiency targets incorporated in price limits. This applies even at the 80% and 90% figures for efficiency content in the residual which Ofwat has proposed.

## **7. Summary and implications**

We have developed a method for assessing the proportion of the reported residual in Ofwat's modelling which can reasonably be attributed to inefficiency. We have then examined, using simulation, the implications of different values for these proportions for the degree of misclassification of companies into efficiency bands.

The method we have used to assess the weight to be given to efficiency does not produce precise estimates. However we believe they are generally of the correct order of magnitude and are more soundly based and more plausible than Ofwat's proposed error allowances. It is hoped that this paper will stimulate further debate on the reliability of the efficiency models.

For the case where the dangers of overestimation and underestimation of potential efficiency improvements are equally balanced we have suggested modifications to Ofwat's procedures which will allow the errors on average to be adjusted for. The procedure also allows the trade offs between operating cost and capital maintenance to be taken into account in a more satisfactory way than at present.

However, our simulations reported in Section 6 suggest that, over a broad range of values for the key parameters, regression models will produce a significant degree of misclassification of companies into efficiency bands. Consideration of the risks that this involves, and the costs and benefits of different approaches to dealing with this misclassification are, however, beyond the scope of this paper.

## Appendix A1: the econometric framework.

An econometric equation is a mathematical statement of a model. All models involve simplifications which abstract from some features of the world which is being modelled. Econometric models are somewhat unusual in that the abstractions themselves form part of the overall model. The following paragraphs inevitably contain some mathematics, to illustrate in concrete terms the components of this abstracting process.

The simplest econometric models involves a dependent variable  $y$ , which in our case is in some function of the cost we wish to model, and an explanatory variable  $x$ , which is the cost driver or explanatory factor. (In what follows we shall assume that  $y$  is either  $\log(\text{cost})$  or  $\log(\text{unit cost})$ .) A naïve version of the cost model is

$$y_i = a + bx_i + u_i \quad (1)$$

Here  $u_i$  might be taken to be two things simultaneously: a measure of relative efficiency and a random error term is taken from a population which is normally distributed with zero mean and constant variance. For unbiasedness (i.e. the estimate of  $b$  is *on average* correct) it is important that  $u_i$  is distributed independently of  $x$ . To a first order approximation  $u_i$  is the proportion by which company  $i$ 's costs are above the regression line (if positive) or below (if negative). (This is a consequence of the assumption that  $y$  is expressed in logs.)<sup>10</sup>

This model omits key sources of inaccuracy in efficiency scores; but it is not itself totally accurate as long as we only have a finite number of observations. Statistical analysis yields an estimate of the important coefficient  $b$ . Different samples taken from the hypothetical population of companies would yield different estimates  $\hat{b}$  of  $b$ . Errors in the estimation of  $b$  will lead directly to errors in the estimation of  $u_i$ , the relative efficiency measure.

In practice there are at least three other sources of error – measurement errors in  $y$ , omitted variables and “errors in variables”. The first is self explanatory, the second relates to the fact that  $x$  will at best only be a measured proxy for the true or ideal cost driver.

### *Measurement errors in $y$*

The existence of joint costs creates a problem of cost allocation. For example, because production takes place through time with durable equipment and stocks of material, until a company is finally wound up it is not possible to say what each of its costs has been. Even with detailed regulatory accounting guidelines in place reasonable people may come to different judgments about particular cash flows and where they belong in the accounts.

---

<sup>10</sup> Strictly the proportion by which the observed cost deviates from the fitted cost is  $\exp(u_i) - 1$ . e.g. if  $u_i = 0.15$  observed costs are 16.18% higher than fitted costs.

For this reason it is accepted that the reported costs underlying the models contain some errors in the  $y$ . To incorporate this fact we need to amend our basic equation as follows:

$$y_i = a + bx_i + \varepsilon_i + u_i \quad (2)$$

Here  $\varepsilon_i$  (epsilon) is the measurement error in  $y$ .

#### *Omitted variables*

Variables are omitted either because there is no suitable measure or because, given the number of observations, they do not pass the test for statistical significance. The latter may occur either because they are small individually or because there is insufficient variation in the sample to pick up identify an effect, or because the peculiarities of this particular sample show an apparent lack of effect. In general we may expect there to be a number of omitted influences. We call each variable  $z^j$  and its marginal effect  $c^j$ :

$$y_i = a + bx_i + \sum_j c^j z_i^j + \varepsilon_i + u_i \quad (3)$$

If each  $c^j z^j$  term is small and if there are several of them then the term behaves exactly like  $\varepsilon_i$  and can be expected to be normally distributed, with a similar effect on the estimates. However some of these individual effects could in principle be quite large. In some cases Ofwat acknowledges the existence of these effects as individual company effects. The onus is on each company to provide persuasive evidence substantiating the importance of any such claimed effects.

In most econometric discussion  $c^j z^j$  and  $\varepsilon_i$  are bundled together and treated as a single item, referred to simply as “the error term.” It is almost unheard of for the variance of the error term to be zero, except where the econometric practitioner has unwittingly tried to estimate an exact accounting identity.

#### *Errors in variables*

The final source of error we shall consider arises because:

- i)  $x$  may itself be subject to measurement error; and
- ii) even if  $x$  is measured accurately, it may be at best a proxy for some concept for which it has not been possible to find a “better” measure. For example, the proportion of water main above 300mm in diameter, which appears as a variable in the water distribution opex model, is now regarded by Ofwat as a proxy for “urbanisation”.

To incorporate this into the model we distinguish between  $x^T$ , the true value of  $x$ , and  $x^M$ , the measured value of  $x$ . The true model is:

$$y_i = a + bx_i^T + \sum_j c^j z_i^j + \varepsilon_i + u_i \quad (4a)$$

but we estimate:

$$y_i = a + bx_i^M + \sum_j c^j z_i^j + \varepsilon_i + u_i \quad (4b)$$

Where  $x_i^M = x_i^T + v_i$ . i.e.

$$y_i = a + bx_i^T + b_i v_i + \sum_j c^j z_i^j + \varepsilon_i + u_i \quad (4c)$$

The error term now has four components, and so will its estimate, which is the residual from the equation. Unlike the other sources of error in the equation however, the presence of the  $bv_i$  term creates bias in the estimation of  $b$ , which arises because the error component  $bv_i$  is correlated with the measured value of the explanatory variable  $x^M$ .

We may express our final model as

$$y_i = a + bx_i + w_i \quad (5)$$

$$\text{Where } w_i = bv_i + \sum_j c^j z_i^j + \varepsilon_i + u_i \quad (6)$$

The residual from the regression is then an estimate of the sum of these individual error terms. If all of these error components are distributed independently of each other then we may write down the overall variance of the error term as the sum of the individual variances. The question of the proportion of the residual attributable to inefficiency is at one level simply a matter of discovering the ratio  $\frac{Var(u)}{Var(w)}$ .

However, the error term is not the same as the residual. A little manipulation of the above equations shows that the residual contains the following terms:

$$r_i = u_i + (a - \hat{a}) + x_i^M(b - \hat{b}) + b(x_i^T - x_i^M) + \sum_j c^j z_i^j + \varepsilon_i \quad (7)$$

Thus relative efficiency is only one of six<sup>11</sup> possible components of the error term:

1. The first term  $u_i$  is the true relative efficiency
2. the difference between the true and estimated constant term *plus*
3. The measured value of  $x$  *times* the difference between the true and estimated values for the slope coefficient
4. The true value of the slope coefficient *times* the difference between the true and measured values of  $x$ .
5. The sum of the omitted effects
6. The error in the measurement of  $y$

In addition, there may be errors due to the use of the wrong mathematical form.

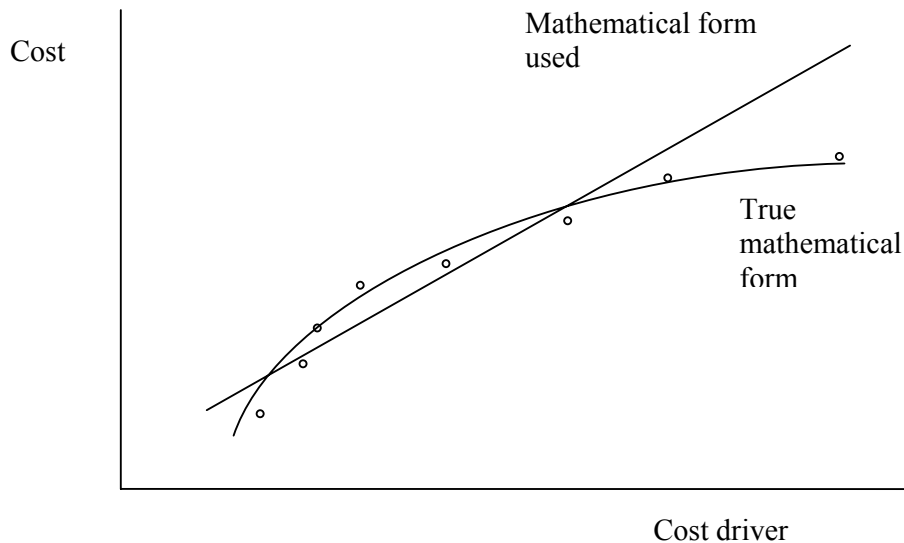
#### *Mathematical form*

In addition to the stochastic errors described above, the mathematical form used may be another source of inaccuracy. The mathematical function employed will almost certainly be a simplification of the underlying true relationship between cost and its

<sup>11</sup> This decomposition is for the simplest case of only one regressor.

drivers since we often have no theory to guide us on the exact form. (See Figure A1.1.)

**Figure A1.1. Effect of using approximate mathematical form**



As the figure indicates, the effect of the approximation will vary. Extreme values may be most subject to distortion. On the other hand, where the cost driver shows very little variation the effect should be fairly small.

(Of these, the error in the constant term cancels out as companies' residuals are compared with each other and can be disregarded.) However, the other items ("equation errors") are all specific to the individual company.

This means that a large positive residual is most likely to contain elements 3-6 which are positive. Conversely, a company with a large negative residual is probably benefiting from negative equation errors.

However, a company which comes average in the rankings *may* appear average because of offsetting items. It could either be inefficient with negative equation errors or efficient with positive equation errors.<sup>12</sup>

This involves detailed consideration of the likely impact of the separate components. But such an approach does not assess the impact of the error on the efficiency measures. In general this will be rather complex.

One way of dealing with this complexity is through simulation. i.e. creating our own data-generating process and looking at the impact of the separate error components on the model's ability to discriminate between efficient and less-efficient companies. We report on this in Section 6.

<sup>12</sup> Stochastic Frontier algorithms tend simply to break down the OLS residual proportionately to estimated variance components and for this reason show a high correlation between OLS and SFA relative efficiency scores, as in the 2001-2 Ofwat efficiency report.

## Appendix A2 Measures of goodness of fit used in the Tables.

The coefficient of determination is often quoted, both by Ofwat and the companies themselves. It measures that proportion of the original variation in the dependent variable which is represented by the fitted values in the regression equation. As a measure of the quality of a cost function it has a number of drawbacks:

- i The amount of the “original variation” in  $y$  depends on whether  $y$  is raw cost, log of cost, or unit cost. The latter will have the smallest original variation so  $R^2$  will tend to be smaller in those equations.
- ii Where raw cost or log cost are used there will be a scale variable which will account for most of the variation in  $y$ .
- iii  $R^2$  does not tell us how much variation in the efficiency estimate there is in order to assess whether it is plausible.

However, for unit cost models  $R^2$  *does* tell us how much of the original variation in unit cost has been attributed to the explanatory factors. If  $R^2$  is very small there is effectively no adjustment for the explanatory factors and the resulting estimated efficiencies will be close to the unadjusted unit measures.

For models involving explicit scale factors, however, a high  $R^2$  does *not* mean that a lot of adjustment has taken place, since most of the  $R^2$  may be due simply to the scale factor.

For these reasons we supplement  $R^2$  with two other measures of fit.

- a) The standard error of estimate (SEE). This is the standard deviation of the residual. where the dependent variable is log cost or unit cost this approximately measures the proportionate variation in unexplained costs.
- b) The standard deviation of the percentage residual. This should be approximately 100 times the SEE and has the advantage that it can be calculated both for the regression and the unit cost models.

We also examine the implied range of efficiency scores as follows.

We first define the adjusted cost index for a firm as its relative expenditure after adjustment for scale and the included explanatory factors, setting the average firm at 100.

We take the ratio of the adjusted cost index of the firm (firm A) with the largest positive percentage residual to the adjusted cost of the firm (B) with the largest negative residual.

e.g. if A has a residual of + 50 % its adjusted cost index is 150. If B has a residual of -50% its cost index is 50. The ratio is then  $150/50 = 3.00$ .

The inverse of the ratio is the implied inefficiency rating of the firm with the highest positive residual, taking the firm with the largest negative residual as the benchmark

In the above example, suppose that as much as 20% of A's cost is over-allocated to the activity in question, and a similar proportion under-allocated to B's activity and that there are no other sources of error (!). A's "true" adjusted cost is 120 (=150-20%) B's "true" adjusted cost is 60 (=50+20%). The "true" cost ratio is then  $120/60 = 2.00$  implying an efficiency level for A of 50% in this activity

However, this question is complicated by the fact that, even in the best models, a proportion of the equation error is almost certainly due to differences in cost allocation. This means that, in general the aggregate model will be better than the individual model since a company which loses on one model will gain on another. This is the "swings and roundabouts" argument, which we discuss further in Section 4.

## Appendix A3: Sewerage non-infrastructure model

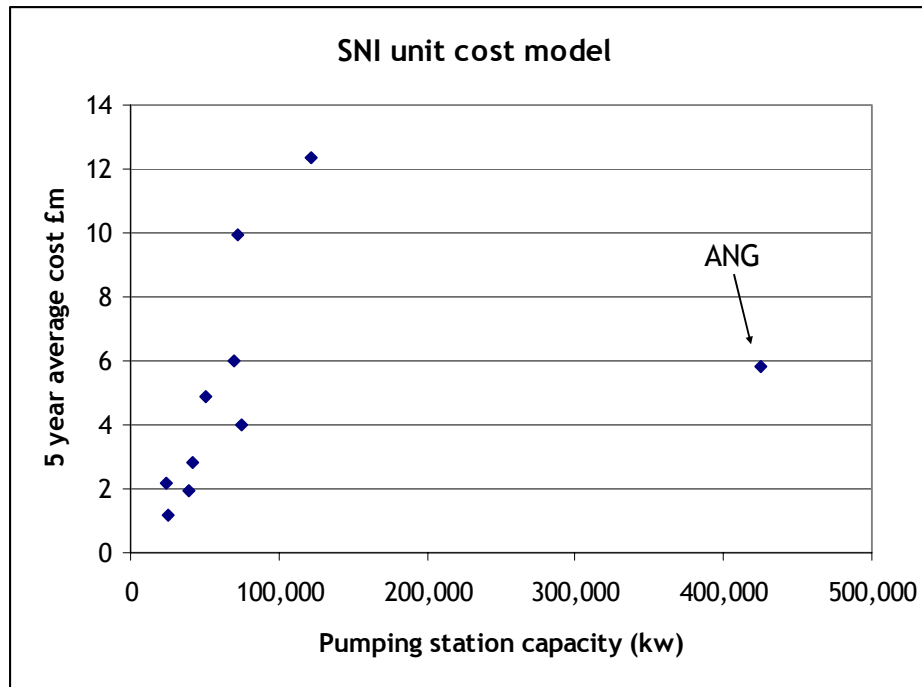


Chart A3. 1. Sewerage non infrastructure model, company totals

