OFWAT

COST ASSESSMENT – ADVANCED ECONOMETRIC MODELS

20 March 2014

FINAL REPORT

Submitted by:

Cambridge Economic Policy Associates Ltd.

CEPA

CONTENTS

IMPORTANT NOTICE

GLOSSARY

Term	Definition
ASHE	Annual Survey of Hours and Earnings
Baseline	A cost value, derived from the model forecasts/company business plan forecasts, which is used in a menu or price control.
BCIS	Building Cost Information Service
Between estimator	Refers to the variation across comparators' explanatory variables in a data set. It is used in conjunction with the within estimator (variation in the company's explanatory variables over time) in panel or pooled regressions to estimate the coefficients on explanatory variables.
Capex	Capital expenditure
Cobb-Douglas model	The Cobb-Douglas (or log-linear) model transforms the variables into logarithms prior to estimation. This model is deemed superior to a linear model in the cost modelling literature as it does not require marginal costs to be constant as in the linear model. Even so, the Cobb-Douglas model is in itself restrictive because, inter alia, it assumes that the extent of returns to scale is the same irrespective of firm size. Compare with translog model.
Corrected OLS (COLS)	See ordinary least squares (OLS) defined below. COLS follows the same statistical technique as OLS (i.e. estimating a line of best fit by minimising the sum of squared errors), however the 'average' line is shifted towards a 'frontier' point i.e., this may be an upper quartile (best) performing company in terms of relatively low costs for its level of outputs. The average line is shifted by changing the intercept point, but no change is made to the slope of the line.
Correlation (coefficient)	A correlation coefficient is the measure of linear interdependence between two variables. The value ranges from -1 to 1, with -1 indicating a perfect negative correlation and 1 indicating a perfect positive correlation. Zero indicates the absence of correlation between the variables.
Corridor	The range calculated by using the model parameters, against which company cost forecasts are evaluated.
Data envelopment analysis (DEA)	A quantitative non-parametric technique that optimises the number of inputs required for a particular output and vice versa. It does not require assumptions on the functional form, but it also does not allow statistical testing on the significance of explanatory variables.
FPL	Future Price Limits
Generalised least squares (GLS)	GLS is a technique for estimating the unknown parameters in a linear regression model. It is applied, for example, when some of the assumptions of the classical regression model break down – such as when the variance of the disturbances is assumed to be non-constant across observations (heteroskedasticity) or when there may be correlation between the disturbances (autocorrelation). The technique is used to estimate the random effects panel model (where there is dependence between

Term	Definition
	observations of the same firm over time).
Hausman test	This test provides information on whether the fixed or random effects treatment is most appropriate. A high value of the statistic (which represents a rejection of the null hypothesis) indicates that the fixed effects model is preferred to the random effects model. Otherwise the random effects treatment is preferred.
Heteroskedasticity	One of the assumptions underpinning the classing linear regression model is that the disturbances are homoskedastic (that is have a constant variance). When the disturbances are heteroskedastic this means that the variance of the disturbances is not constant across firms (an example is where the disturbances increase as firm size increases).
I&C	Industrial and commercial customers
IRC	Infrastructure renewals charge (annual allowance)
IRE	Infrastructure renewal expenditure (actual)
Maximum likelihood estimation (MLE)	This is a method of estimating the parameters of a statistical model. Under the standard assumptions underpinning the classical linear regression model, MLE produces identical estimates to those produced by OLS. However, MLE has been shown to have desirable (large sample) properties under a wide range of assumptions (unlike OLS) and this method is therefore used in a wide range of contexts, including stochastic frontier analysis. Information is needed concerning the distribution of the errors to implement MLE.
Menu regulation	Menu regulation is a form of regulation where regulated companies are no longer presented with a 'take it or appeal it' regulatory offer regarding the allowed level of expenditure, but are instead given a range of options from which to choose.
MNI	Maintenance of non-infrastructure expenditure (actual)
Multicollinearity	An exact linear relationship between two or more explanatory variables characterises the extreme case of perfect collinearity (approximate linear relationships between variables are more common in practice). In the former case (perfect collinearity) the OLS procedure cannot be implemented. The latter case (approximate linear relationships) results in high standard errors. Whilst the parameter estimates and estimates of the standard errors are not biased as such, the problem is that it will be hard to draw conclusions on the impact of individual variables on the dependent variable. The overall predictive power of the model is not reduced (only the ability to use the coefficients individually).
Opex	Operating expenditure
Ordinary Least Squares (OLS)	OLS is a method by which linear regression analysis seeks to derive a relationship between company performance and characteristics of the production process. This method is used when companies have relatively similar inputs and outputs. Using available information to estimate a line of best fit (by minimising the sum of squared errors) the average cost or production function is calculated.

Term	Definition
Pooled OLS	The pooled OLS model treats the data as if it was a cross-section – that is, e.g. 90 firms, rather than a panel of 10 firms over nine years. This approach does not therefore recognise the panel structure of the data, and can be tested against the panel model variants. It is however a simple model that is used by economic regulators in particular.
Pooled Stochastic Frontier Analysis (SFA) model	This is a maximum likelihood estimation model that is the same as COLS except that a one-sided error term is included to permit the existence of inefficiency (with the error term decomposed into its noise and inefficiency components). This approach requires distributional assumptions on the error components.
PR14	Price Review 2014
Real price effects (RPEs)	The amount by which certain input prices are expected to move relative to RPI (either increased/ decreasing at a faster rate).
Regional BCIS index	A proxy for regional differences in construction prices, based on tender prices from the BCIS.
Time invariant efficiency model: Fixed Effects (FE)	This is the standard fixed effects model used in the panel data literature, except that in this case the fixed effects terms are given an inefficiency interpretation. In the fixed effects model, firm-specific effects (unobserved differences between firms) are estimated as fixed parameters to be estimated, by including firm-specific dummy variables in the regression. However, the true distinction between fixed and random effects is whether the effects are correlated with the other regressors or not (in the case of random effects the effects are assumed to be uncorrelated with the regressors, whereas in fixed effects the effects are permitted to be correlated with the regressors).
	It is sometimes said that this approach is concerned only with the particular firms in the sample (i.e. that the sample contains all relevant firms and there are therefore no additional firms outside the sample of interest). The random effects model treats the unobserved firm effects as randomly distributed across firms (so here we see the current sample as being drawn from a wider sample or population). It has been pointed out in the literature that in fact the fixed effects model can be reformulated and estimated as a random effects model, so the distinction concerning whether the effects are stochastic or not is erroneous (see, for example, Greene, Econometric Analysis, 5th Edition, page 285).
Time invariant efficiency model: Random Effects (RE)	This is the standard random effects model used in the panel data literature, except that in this case the random effects terms are given an inefficiency interpretation. The random effects specification imposes the assumption that the unobserved individual effects are uncorrelated with the regressors.
Time-invariant SFA model	This is a maximum likelihood model and an extension of the random effects model but now with distributional assumptions imposed and with estimation proceeding via MLE, not generalised least squares (GLS), as in the standard panel data random effects model. See Pitt and Lee (1981).

Term	Definition
Time varying SFA model	This is a maximum likelihood model that extends the model above to permit efficiency to vary over time but in a restricted way, since the direction of efficiency change over time must be the same for all firms (and thus rankings cannot change). See Battese and Coelli (1992)
Skewness	Skewness is a term used to describe non-symmetric distribution (a right skewed distribution has a longer "tail" to the right and vice versa for a left skewed distribution).
STW	Sewage treatment works
Total factor productivity (TFP)	A measure of the economy's long-term technological change.
Totex	Total expenditure (opex + capex)
Translog model	The translog model is one of the so-called flexible functional forms and is used routinely in the academic literature. In the current context one of its particular advantages is that it allows the degree of returns to scale to vary with firm size. The Cobb-Douglas is nested within the translog so it is possible to test the Cobb-Douglas restriction.
Triangulation	The use of multiple methodologies and the numbers from them (averages, max, min etc.) to come up with a single value for cost assessment.
UKWIR	UK Water Industry Research
WaSC	Water and sewerage company
Within estimator	Refers to the variation in the company's explanatory variables over time in a data set. It is used in conjunction with the between estimator (variation across companies' explanatory variables) in panel or pooled regressions to estimate the coefficients on explanatory variables.
WoC	Water only company
WTW	Water treatment works

EXECUTIVE SUMMARY

Introduction

Since August 2012 CEPA, in conjunction with Dr Andrew Smith of the University of Leeds, has been assisting Ofwat in developing water and sewerage econometric cost models. In January 2013 Ofwat published CEPA's *Cost Assessment Report*[1] as part of their methodology consultation, which discussed the viability of totex modelling in water and sewerage. Since the January report we have received new data from Ofwat, the August 2013 data, and have used this to retest and refine a broad range of models. The models presented in this report use the most recent data, spanning up to 2012-13. They cover total expenditure (totex) in wholesale water and base expenditure (operating and base service capital maintenance expenditure) in wholesale sewerage. Ofwat has modelled sewerage enhancement separately, mainly using unit cost models. In agreement with Ofwat, we excluded several types of costs from the econometric modelling – such as third party costs – as those are beyond the companies' control. Ofwat are addressing these costs separately in the risk-based review.

A report prepared by Jacobs on behalf of Ofwat will be published alongside this report. The Jacobs' report sets out forecasts for the explanatory variables used with the recommended models to help Ofwat set the cost benchmarks for the companies.

Table E.1 below provides a summary of the cost areas included in the advanced econometric models. We model different expenditure breakdowns in water and sewerage.

Table E.1: Expenditure modelled

Type of expenditure	Water	Sewerage		
	Wholesale	Wholesale	Network	Treatment & sludge
Opex + base capex	✓	✓	✓	✓
Totex	✓	✗	✗	✗

In water, we have some models that cover all of totex, while others only cover base expenditure, i.e. excluding enhancement capex.[2] In sewerage, we approached modelling in a slightly different way. We attempted to model totex but it did not prove viable. Therefore, all the sewerage models presented in this report exclude enhancement capex. The data allowed us to split costs between network and treatment/ sludge, and, to model these areas separately as well as modelling them together as wholesale base sewerage expenditure.

We worked with Dr Andrew Smith and Ofwat to develop the models to use for calculating cost allowances at PR14 and then to test their robustness. This process began in August 2012 and our development has included an initial consultation with UKWIR and specific inputs on technical issues from several academic advisors. We recognise that given the data constraints and a range of estimation techniques, no econometric model will perfectly reflect all of the

---

[1] CEPA. *Ofwat: cost assessment*. January 2013.

[2] In these cases unit costs are added to determine totex.

companies' characteristics.[3] As such, our proposed approach for Ofwat is for them to use a number of models with different variables and/or estimation techniques, and triangulate between these models to determine robust cost benchmarks for the companies. We have tested the modelling and undertaken external Quality Assurance (QA) – as a result, we consider that our analysis and recommendations are in line with regulatory best practice.

Model selection

Our model selection process began with viability testing of totex and opex plus base capex models. When we established that modelling was viable we received additional and revised data from Ofwat covering the years up until 2012/13 – the August 2013 data-set. In order to choose between models, five standard and commonly implemented criteria were used to assess a long list of models:

- theoretical correctness;

- statistical performance;

- practical implementation issues;

- robustness testing; and

- regulatory best practice.

We used these criteria to first reduce our long list of models and then refined this list further by focusing on the statistical performance and robustness testing criteria. We found it difficult to identify suitable metrics to help choose between models in a mechanistic way, so we have adopted an approach based on a 'traffic-light' system to indicate how well the model performs against a given criterion, i.e., a 'green light' corresponds to 'good', 'amber light' corresponds to 'acceptable but with a few issues', and a 'red light' means that the model is flawed.

We did not assign a red light to any model for theoretical correctness as the models had already been narrowed down to a theoretically robust set in discussions with Ofwat, UKWIR and by implementing established econometric approaches to modelling. The other categories – statistical performance and robustness testing – do allow for a red traffic light, in which case the model would no longer be considered a candidate. For the former, a red light indicates that several of the core parameter estimates are substantially outside our expectations. For robustness testing it means that either the efficiency scores resulting from the model or the prediction are implausible; or that there is significant evidence for having different coefficients in different time periods.

Our final selection process is summarised in Figure E.1.

---

[3] We note that we would not expect any of the models to perfectly predict companies' expenditure due to inefficiencies.

Figure E.1: Model Selection Process
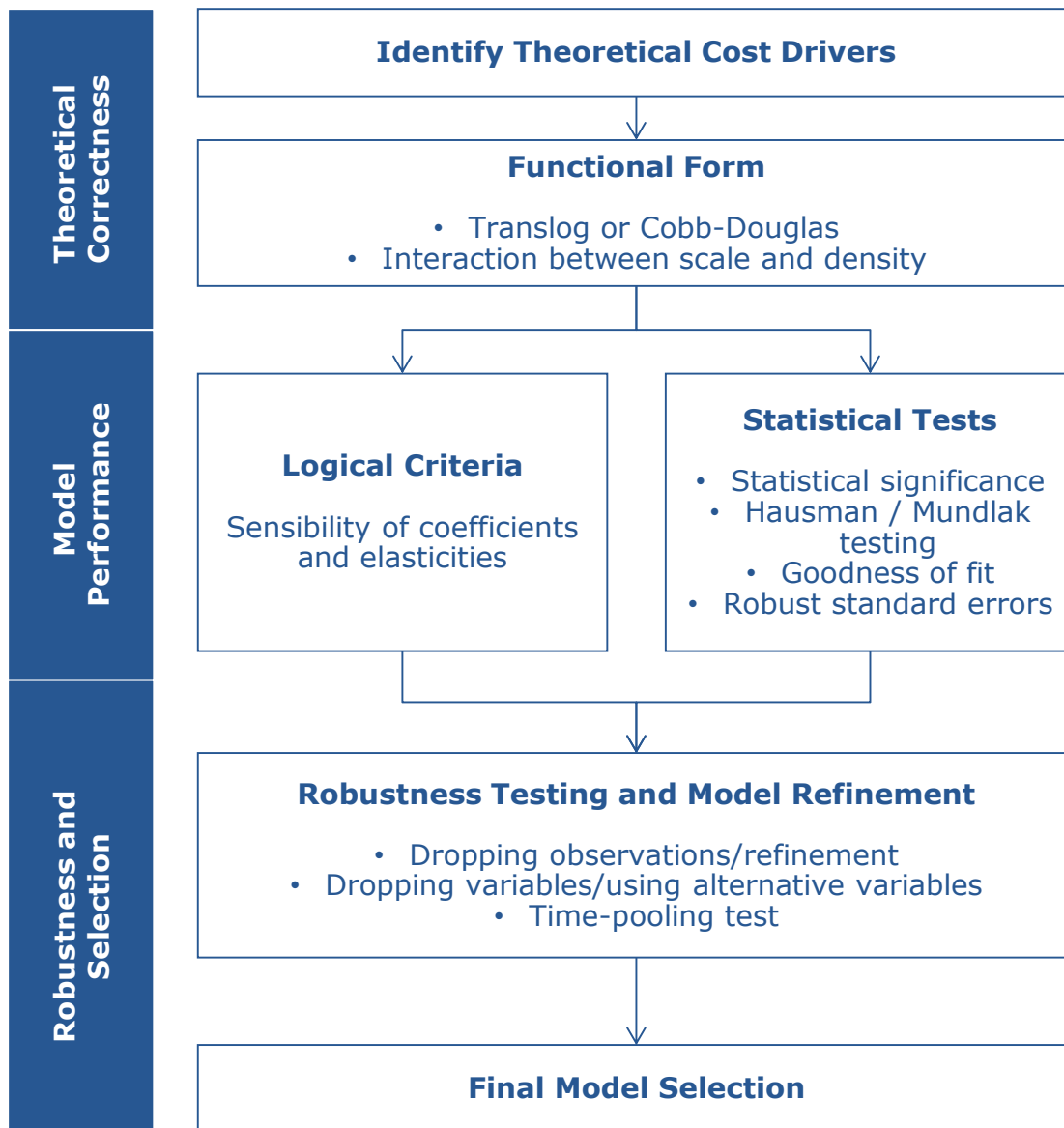
We believe the preferred models provide a range of efficiency specification methods (time-invariant efficiency and time-varying), estimation techniques (GLS [RE] and OLS), and full and refined models where available.

All our preferred models are in log form (which means the coefficients can be interpreted as elasticities) and allow for different economies of scale for different size companies (referred to as translog models). Our testing and other studies in this sector supported this choice.[4] While these types of models are less transparent than standard non-varying economies of scale (which we refer to as Cobb-Douglas [CD]) specifications they better reflect the reality of the economies of

---

[4] For example see Stone & Webster, *Investigation into evidence for economies of scale in the water and sewerage industry in England and Wales: Final Report*, prepared for and published by Ofwat, 2004, and Saal et al, *Scale and scope economies and the efficient configuration of the water industry: a survey of the literature*, Aston Centre for Critical Infrastructure and Services Working Paper, Aston University, UK, 2011.

scale present in the water and sewerage industry.[5] Our use of a log specification does mean that the cost predictions generated may be biased, either over- or under-estimated depending on the shape of the production function, and an adjustment factor is required to ensure that the linear transformation of the cost predictions are not biased.[6] We have proposed that Ofwat use an adjustment factor in line with that used by Ofgem for DPCR5 and RIIO-GD1. Ofgem referred to this as the 'alpha correction factor'.[7]

All the models selected excluded regional BCIS as there is a high correlation between this variable and the regional wage variable. We found that models that included BCIS resulted in unexpected coefficients. We believe that the regional wage variable explains more of the regional price variations than the BCIS.

We recommended that Ofwat use five water models and five sewerage models. For water we proposed three model specifications, run using GLS (RE) and/or OLS. Our recommended water model specifications were:

- A **full model** specification including all explanatory variables provided to us by Ofwat including our estimation of the regional wage variable, but excluding BCIS. Model WM3.[8]

- A **refined model** specification including only variables which we found to be statistically significant or were important cost drivers from a theoretical perspective. Models WM5 and WM6.

- An **opex plus base capex model** using similar explanatory variables to the refined model above, but excluding enhancement expenditure. Models WM9 and WM10. Ofwat modelled the enhancement expenditure separately.

Table E.2 below lists the preferred water models' performance against the selection criteria. Note, when comparing the models we used on the COLS and GLS (RE) efficiency scores, however when the models are triangulated (discussed later) the efficiency target relies only on a correction factor. Based on this application of the modelling results the only difference between the COLS and GLS (RE) is the weight given to the within (variation over time for a company) and between (variation across companies) estimators, with GLS placing more weight on the within estimators than OLS.

---

[5] Cobb-Douglas is a production function rather than a cost function. We are modelling the latter, but we have however used the term CD as the concept is similar.

[6] This is explained in statistics as Jensen's inequality.

[7] Ofgem, *RIIO-GD1: Initial proposals – Step-by-step guide for the cost efficiency assessment methodology,* August 2012, page 12 and Ofgem, *Electricity distribution price control review; Final proposals – allowed revenue – cost assessment appendix,* December 2009, page 87.

[8] We tested a GLS (RE) fully specified model, however as the number of explanatory variables exceeded the number of companies the between estimator could not be computed. The programme we used, LIMDEP, still estimates the full model, but we do not have confidence in the results produced.

Table E.2: Final water models

	Theoretical correctness	Statistical performance	Robustness check
Totex			
WM3 – full translog COLS without BCIS	G	A	A
WM5 – refined translog COLS without BCIS	G	G	G
WM6 - refined translog GLS (RE) without BCIS	G	G	G
Opex + base capex			
WM9 – refined translog COLS without BCIS	G	A	G
WM10 – refined translog GLS (RE) without BCIS	G	G	G

The sewerage models we selected were all opex plus base capex models. We could not establish a viable sewerage totex model which produced consistent and robust results. Our recommended sewerage model specifications were:

- A **sewage treatment model** specification run with both GLS (RE) and OLS. The explanatory variables were 'refined', as a 'fully' specified model did not produce significantly different results from the refined model. Given that there are only 10 comparators in sewerage we considered the greater number of degrees of freedom gained outweighed any potential small loss of explanatory power. Models SM5 and SM6.

- A **sewer network model** specification run using only GLS (RE). Again we used a refined model as we did not find any advantages from using a 'fully' specified model. We did not use an OLS model as the coefficients were not in line with our expectations and their interpretation was not consistent with those of the cost drivers. Model SM1.

- A **sewerage opex plus base capex model** specification run with both GLS (RE) and OLS. This model specification used similar explanatory variables to the treatment and network models, however as treatment makes up a greater proportion of expenditure the load explanatory variable was preferred to the length variable. Models SM9 and SM10.

- In all cases Ofwat modelled the **enhancement** expenditure separately.

Table E.3 below lists the preferred sewerage models' performance against the selection criteria.

Table E.3: Final sewerage models

	Theoretical correctness	Statistical performance	Robustness check
Network opex + base capex			
SM1 - refined translog GLS (RE)	G	G	G
Treatment & sludge opex + base capex			
SM5 - refined translog GLS (RE)	G	G	G
SM6 - refined translog COLS	G	G	G
Wholesale opex + base capex			
SM9 - refined translog GLS (RE)	G	G	A
SM10 - refined translog COLS	G	G	A

Triangulation and efficiency estimation

As we had recommended the use of multiple models to Ofwat, an approach to establish a single estimate across these models was required, for water and sewerage in turn, i.e. a triangulation method. Our proposed triangulation method was based around the following criteria:

- maximising the intermediate information each option offers, i.e. estimate from 'bottom-up' models capturing different parts of the value chain and estimate from 'top-down' models capturing the whole value chain;

- transparency;

- logical flow, i.e. do the weights placed on each model make intuitive sense; and

- ease of implementation/ replicability.

Our recommended approach follows a logical process of estimating separate elements of the value chain or cost categories (we term these bottom-up models) and top-down models (capturing the whole value chain/ more aggregated costs) before triangulating these together to get a single prediction. Based on this approach and given the need to avoid 'cherry-picking' results (i.e. selecting the upper quartile in all models),[9] we recommended that the calculation of the cost benchmarks be done based on the final single prediction.

We recommended that the simple ratio approach to estimating efficiency should be used.[10] This is a transparent approach which avoids cherry-picking, is replicable and has regulatory precedent (this and alternative approaches are discussed in section 5.2).[11]

---

[9] When we refer to the upper quartile we are referring to the upper quartile efficiency performance, which is equivalent to a lower quartile cost.

[10] Rather than using both forms of efficiency estimation (e.g. based on residuals from the econometric modelling and ratio).

[11] Ofwat used ratios in PR09 and Ofgem has used ratios for RIIO-GD1 and RIIO-ED1 fast track decisions.

We note that we found small differences between the alternative options of triangulating at different stages of the modelling, or using a mix of residual and ratio efficiency estimation.

In addition, we recommended to Ofwat that the efficiency adjustment be calculated on historical data rather than forecast expenditure. Using historical data means that the companies are compared against the relative past performances rather than their future estimated performance. In the former case, there would be no limit on the number of companies which could be determined as 'upper quartile' performers against the benchmark. If the forecast expenditure was used, then there would a limited number of 'good' performers as there are a fixed number of companies in each quartile.

We did not provided a recommendation to Ofwat on how far from the average industry performance they should set the cost benchmark, e.g., upper quartile/ upper third. We do however consider that this should be based on the level of confidence Ofwat has in the predictions from the modelling and how challenging they wish to make the targets for the companies. This will be a matter of regulatory judgement by Ofwat.

1. INTRODUCTION

Since August 2012 CEPA, in conjunction with Dr Andrew Smith of the University of Leeds, has been assisting Ofwat in developing water and sewerage econometric cost models. In January 2013 Ofwat published CEPA's *Cost Assessment Report* as part of their methodology consultation, which discussed the viability of totex modelling in water and sewerage. Since the January report we have received new data from Ofwat, the August 2013 data, and have used this to retest and refine a broad range of models. The models presented in this report used the most recent data, spanning up to 2012-13. They cover total expenditure (totex) in wholesale water and base expenditure (operating [opex] and base service capital maintenance expenditure [capex]) in wholesale sewerage. Ofwat has modelled sewerage enhancement separately, mainly using unit cost models. These Ofwat models are discussed in a separate report published alongside this one.

A report prepared by Jacobs on behalf of Ofwat sets out forecasts for the explanatory variables used with the recommended models to help Ofwat set the cost benchmarks for the companies.

1.1. Objective

This report sets out the testing that we undertook to get to a set of robust models for water and sewerage. It also sets out our recommendations for assessing costs for these services in PR14. We worked alongside Ofwat to ensure the modelling is consistent with the rest of the PR14 framework. We also shared initial results of our totex models with the UKWIR steering group in September 2012 to better understand what the industry viewed as its main cost drivers. This also allowed us to understand and build-on the total expenditure benchmarking work undertaken by Reckon on behalf of UKWIR.[12]

Dr Andrew Smith, of the University of Leeds, took a leading role in the initial development of the approach and definition of possible model structures. He then continued to provide support and guidance to the CEPA team during the testing of various models and the determination of preferred options. This included the provision of expert advice and guidance during the robustness testing phase of the project.

In addition, Dr Michael Pollitt, of the Judge Business School at the University of Cambridge, and Jon Stern, of the Centre for Competition and Regulatory Policy at City University, have provided independent external review of the approach we have adopted. We also sought technical advice from Professor William Greene, of the NYU Stern Business School, on the principles of random effects versus corrected ordinary least squares and separating unobserved heterogeneity from inefficiency. They have not reviewed the final models we assess in this report but we have taken their comments into account when selecting the preferred set of models.

1.2. Changes since the January 2013 'CEPA Cost Assessment Report'

Since the publication of the CEPA *Cost Assessment Report*, we have conducted additional modelling and updated our analysis using the latest dataset which included the companies' August 2013 submissions. There are several significant changes to the results presented in the CEPA *Cost Assessment Report*, namely:

---

[12] UKWIR, *A total expenditure approach to cost assessment,* 2012, http://www.ukwir.org/web/ukwirlibrary/95954.

- We are no longer modelling sewerage opex at a sub-company level. Instead, we prefer the use models that combine opex and capex to avoid capex bias.

- We were also able to model treatment base capex and sludge base expenditure (opex and maintenance capex) due to revisited data splits.

This led to an increase in the coverage of the econometric modelling, which in turn has reduced the use of unit cost models. In agreement with Ofwat, we excluded several types of costs from the econometric modelling – such as third party cost – as those are materially uncertain. Ofwat are addressing these costs separately in the risk-based review. Table 1.1 below provides a summary of the cost areas included in the advanced econometric models. We model different expenditure breakdowns in water and sewerage.

Table 1.1: Expenditure modelled

Type of expenditure	Water	Sewerage		
	Wholesale	Wholesale	Network	Treatment & sludge
Opex + base capex	✓	✓	✓	✓
Totex	✓	✗	✗	✗

In water, we have some models that cover all totex, while others only cover base expenditure, i.e. excluding enhancement capex.[13] In sewerage, we approached modelling in a slightly different way. We attempted to model totex but it did not prove viable as indicated in the CEPA Cost Assessment Report. Therefore, all the sewerage models presented in this report exclude enhancement capex. The data allowed us to split costs between network and treatment/sludge, and, to model these areas separately as well as modelling them together as wholesale base sewerage expenditure.

1.3. Process

The process we have followed in developing the econometric cost assessment models is set out in Figure 1.1 overleaf. This process included quality assurance via ongoing discussions with Ofwat, as well as input from UKWIR and technical advice from academic experts.

As discussed above, the introduction of the August 2013 data meant that we had to revisit the viability of the models before we decided on a long list to assess.

---

[13] In these cases unit costs are to determine totex.

Figure 1.1: Model development process
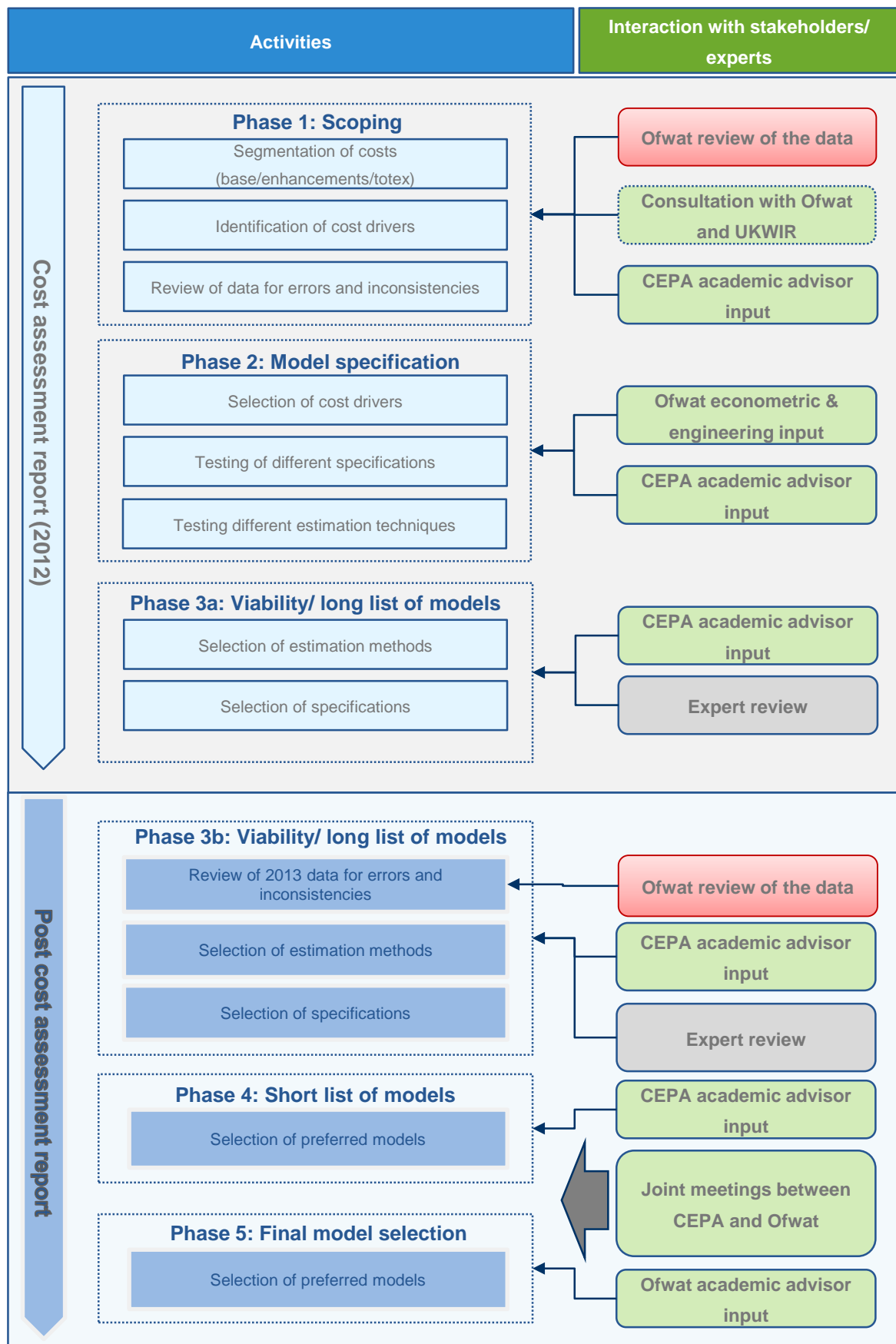
It should be noted that the CEPA Academic Advisor, Dr Andrew Smith, was appointed as the Ofwat Academic Advisor mentioned in the figure above during the model development process.

1.4.   Structure of the report

The report continues as follows:

- Section 2 describes our approach to modelling and the main issues we have looked at while testing, such as explanatory variables, economies of scale, efficiency assumptions, capex smoothing and panel length;

- Section 3 sets out the criteria we have used to assess each viable model, including our scoring system;

- Section 4 presents the preferred water and sewerage models; and

- Section 5 discusses triangulation options and efficiency adjustments.

The report also includes a number of annexes which give more detail on the testing we have done and alternatives considered:

- Annex 1 sets out the variables used in water and sewerage;

- Annex 2 discusses alternative variables that we have considered or tested;

- Annex 3 describes how we constructed the regional wage variable used in the final models;

- Annex 4 presents the detailed results for a selection of the water models;

- Annex 5 presents the detailed results for a selection of the sewerage models;

- Annex 6 details the efficiency calculations and adjustments, associated with different types of estimators;

- Annex 7 details the options for transforming logarithmic values into level values;

- Annex 8 provides the non-normalised coefficients for the final models recommended; and

- Annex 9 provides recommendations for cost modelling in PR19.

2. APPROACH TO MODELLING

As part of our analysis we have tested a wide range of models using the latest dataset, updated after the August submission, consistent with the cost drivers, methods and functional forms that we used during the previous stages of the analysis. These included translogs versus Cobb-Douglas (CD) functional forms (discussed in Section 2.2); ordinary least squares (OLS), generalised least squares (GLS) random effects (RE), fixed (FE), stochastic frontier analysis (SFA), and true random effect estimations (discussed in Section 2.3); the choice of panel length (discussed in Section 2.4); and smoothed versus unsmoothed capex (discussed in Section 2.5).

As we mentioned in the introduction, the data used in our modelling had changed since the publication of our *Cost Assessment Report*. We were also able to add two years of data to the dataset that we started with in August 2012, which meant that the dataset used for the modelling in this report covered the period up to 2012-13.

We note that the final dataset that we used had undergone significant changes, even in the historical costs, as some companies resubmitted their figures. The revisions to the historical data were not consistent across companies in terms of magnitude and direction. This led to changes in the models' coefficients from our earlier cost modelling. We used the companies' expenditure data submitted as part of the June Returns and August submissions as the dependent variable. As noted earlier Ofwat adjusted the historical expenditure to exclude certain wholesale costs that are materially uncertain (e.g. costs associated with third party services).

We discuss the explanatory variables (cost drivers) and then the assumptions, and associated implications, in turn below.

2.1. Explanatory variables

The majority of the variables that we included in our final models are defined in the same way as those we presented in the CEPA *Cost Assessment Report*. However, we tested a number of new variables and redefined a few of the existing variables used previously. In Annex 1 we provide detail on the specification for each explanatory variable and rationale behind their use. Annex 2 discusses alternative variables we considered and our rationale for not using them.

Table 2.1 below presents all the explanatory variables we have tested in the various water models.

Table 2.1: Range of explanatory variables in water models

Type	Variable
Core	Length of mains
	Property density
	Usage
	Time trend
Input prices	Average regional wage
	Regional BCIS index
Network characteristics	Population density (occupancy)

Type	Variable
	Proportion of metered properties
	Proportion of usage by metered household properties
	Proportion of usage by metered non-household properties
Treatment and sources characteristics	Sources
	Pumping head
	Proportion of water input from river abstractions
	Proportion of water input from reservoirs
Activity	Proportion of new meters
	Proportion of new mains
	Proportion of mains relined and renewed
Quality	Properties below reference pressure level
	Leakage
	Properties affected by unplanned interruptions > 3 hrs
	Properties affected by planned interruptions > 3 hrs

While we discuss the variables in more detail in Annex 1, it should be noted that the average wage variable we used is different from that constructed by Ofwat for PR09 and from that used in the January 2013 *Cost Assessment Report*. A brief description of the new variable is set out in Text Box 2.1 below.

Text Box 2.1: Average regional wage

The wage variable has been constructed by CEPA, supported by Ofwat, and is different from the way Ofwat constructed regional wages in PR09. It is based on regional rather than local area wage differences as we consider companies are not restricted to sourcing workforce from the county/area of operation. The variable excludes overtime pay and focuses on hourly rather than weekly pay to eliminate any differences that could be attributed to inefficiency or company policy. In this way, the wage variable is exogenous of the particular company and captures the ability of companies to source labour from areas with different wage profiles. We discuss the construction of this variable in more detail in Annex 3.

As we decided to no longer conduct sewerage modelling at the sub-company level we did not include any drivers at the sub-company level. We did however include additional drivers for treatment and sludge. Table 2.2 below presents all the explanatory variables we tested in the various sewerage models.

Table 2.2: Range of explanatory variables in sewerage models

Type	Variable
Core	Length of sewers
	Density
	Usage
	Time trend

Type	Variable
Input prices	Average regional wage
	Regional BCIS index
Network activity	Proportion of sewers replaced and renewed
Treatment and sludge	Load
	Sludge disposed
	Proportion of load in treatment works size bands 1-3
	Proportion of load in treatment works size bands 4 and 5
	Proportion of loaded treated by activated sludge treatment
	Number of large works with the tight consents dummy

We note that across both water and sewerage models a number of variables were highly correlated with each other (either negatively or positively). We have set out the correlation matrices for the water and sewerage explanatory variables in Annex 1. We discuss the implications of multicollinearity in Section 3.2.1.

2.2. Economies of scale (Cobb-Douglas versus translog)

CD is a production function (which by duality, can be expressed as a cost function) which places weights on the input factors. The CD is a standard functional form used in cost assessment literature. When in a log-linear form the CD allows for the marginal costs to vary and coefficients to be interpreted as the elasticity of cost with respect to the corresponding driver. A translog introduces further flexibility by allowing the economies of scale to vary as well.[14]

We tested both functional forms in our modelling as previous literature indicated that there is evidence of varying economies of scale in the water and sewerage industry. For example, work commissioned and published by Ofwat (Stone and Webster 2004),[15] suggested the presence of variable returns in the water industry, with evidence of diseconomies of scale for water and sewerage companies (WaSCs), but possible economies of scale for WoCs. Although, Stone and Webster could not reject the presence of constant returns to scale for water-only companies (WoCs). In addition, Saal et al (2011)[16] found that, for WoCs, the average sample firm was subject to diseconomies of scale. However, it concluded that vertically integrated firms gained significant benefits from economies of scope and scale. We discussed the theoretical implications of the translog with Ofwat staff and we agreed with them that a translog form was viable.

The results of our testing, using joint statistical significance of the translog terms, consistently showed that translog models were statistically preferred for both water and sewerage.

---

[14] In practice this is achieved by adding the square and cross terms of the main scale variables to the equation.
[15] Supra N4.
[16] Supra N4.

2.3. Estimation methods and efficiency specifications

2.3.1. Range of estimation techniques tested

There are numerous econometric estimation approaches that can be used with panel or pooled data. (The main difference between panel and pooled datasets is that pooled treats all observations as independent while panel data treats companies' observations as being related over time.)[17] As part of our earlier report and this subsequent refinement, we tested a number of approaches. These are set out in Table 2.3 below.

Table 2.3: Estimation methods

Estimation Method	Description
Pooled Ordinary Least Squares (OLS)	The pooled OLS model treats the data as if it was a cross-section – that is, e.g. 90 firms, rather than a panel of 10 water and sewerage firms over nine years. Not recognizing the structure of the data causes the OLS estimator to place equal weight on the between variation (i.e. differences between companies) and within variation (i.e. differences between years for the same company) when calculating the estimate. OLS does not distinguish between white noise, heterogeneity and inefficiency, unlike the rest of the methods which make some assumptions about the decomposition of residuals into noise and other components such as inefficiency.
	Efficiency is calculated in each year using the difference between each firm's residual and the minimum residual for that year (note, different companies may be at the frontier in each year). These efficiencies are then averaged over time (e.g. five years). Although efficiency is allowed to vary over time, we note that there is no structure to this variation. We do not use these efficiency scores in making the efficiency adjustments, however, so these differences are not crucial to the modelling.
Pooled Stochastic Frontier Analysis (SFA)	This is a maximum likelihood estimation (MLE) model requiring distributional assumptions on the error term and is the same as OLS except that a one-sided error term is included to permit the existence of inefficiency (with the error term decomposed into its noise and inefficiency components). This model attempts to distinguish between white noise and inefficiency, but does not try to control for company heterogeneity. The pooled element of this technique means that the data is (like Pooled OLS above) treated as a cross-section, thus the structure of the data is ignored and the same implications follow.
Time invariant panel method - Random Effects (RE)	Panel methods in general have the advantage that estimation takes into account the structure of the data. That is, it recognizes that we have 18 water companies over time, rather than different companies each year. In our case, it uses generalised least squares (GLS), which places more weight on the within variation than OLS when calculating parameter estimates. There are two broad categories of panel methods, RE and FE.
	RE require that firm-specific effects be uncorrelated with cost drivers. The error term thus captures the company effect and white noise. The company effect is assumed to be randomly distributed across firms (within and out of sample). While noise is assumed to have an expected value of zero, thus allowing us to estimate the average company effect, which is interpreted as inefficiency. Efficiency is thus assumed to be constant over time. The model does not distinguish between unobserved heterogeneity and inefficiency.

---

[17] See Section 3.2.3 of the January 2013 CEPA *Cost Assessment Report*.

Estimation Method	Description
	RE models are perceived to yield more precise coefficients than FE and OLS models but have unclear properties in small samples.
Time invariant panel method - Fixed Effects (FE)	FE is estimated via OLS. It allows for company specific effects to be correlated with cost drivers by estimating the company effect as a parameter in estimation (this can then be recast and interpreted as inefficiency). Efficiency is assumed to be constant over time. The advantage of the FE model is that it produces unbiased and consistent parameter estimates in the presence of correlation between company effects and cost drivers. However, these estimates may be less precise than RE estimates. That is, although FE may be unbiased, the point estimates in a particular sample may be less accurate than RE estimates. Other disadvantages of this model include that it cannot deal with time invariant regressors and the inclusion of company effects means that the number of parameters estimated grows with the number of companies.
Time varying true RE	This is a maximum likelihood variant of the above RE model that attempts to decompose the company effect into inefficiency and unobserved heterogeneity. This model assumes that heterogeneity is constant over time while inefficiency can vary. It also requires distributional assumptions about the error and heterogeneity terms. However, this model can have difficulties separating persistent inefficiency from time invariant heterogeneity.
Time invariant panel SFA (Pitt and Lee)[18]	This is a MLE model requiring distributional assumptions on both the error and inefficiency terms. It takes the data structure into account. It is an extension of the RE model but with distributional assumptions imposed on the error and company effects (but doesn't attempt to control for heterogeneity). Estimation proceeding via MLE. For this model, inefficiency is assumed to be constant over time.
Time varying SFA (BC92)[19]	This is a MLE model requiring distributional assumptions on both the error term and on efficiency. It extends the model above (Pitt and Lee) to permit efficiency to vary over time but in a restricted way, since the direction of efficiency change over time must be the same for all firms (and thus rankings cannot change).
Time varying SFA (Cuesta 2000)[20]	This is a flexible version of BC92 (also using MLE estimator) that allows for firm-specific paths of inefficiency. That is, some companies can be catching up or falling away from the frontier in any given year. This model was used by ORR for PR08.
Time varying pooled OLS (CSS)[21]	This model permits firm specific time paths for inefficiency and tries to differentiate between statistical noise and inefficiency (as opposed to pooled OLS that does not differentiate), but without the need to impose distributional assumptions. One disadvantage of the Cornwell, Schmidt and Sickles (CSS) model is that it does not allow us to test the statistical significance of the time variation in inefficiency.

---

[18] See Pitt and Lee, *The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry*, Journal of Development Economics, 9, 43-64. (1981).

[19] See Battese and Coelli, *Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India,* Journal of Productivity Analysis, 3, 153-169. (1992).

[20] See Cuesta R.A. *A Production Model With Firm Specific Temporal Variation in Technical Inefficiency: With Application to Spanish Dairy Farms*, Journal of Productivity Analysis 13 (2): 139-158. (2000).

[21] See Cornwell, Christopher & Schmidt, Peter & Sickles, Robin C., *Production Frontiers With Cross-Sectional And Time-Series Variation In Efficiency Levels*, Journal of Econometrics, 46, 185-200, (1990).

In general, we found that GLS (RE) models were preferred to FE, and that GLS (RE) and pooled OLS models provided more stable and robust results than SFA models.

There are two key differences between a COLS approach using pooled data and a panel RE approach:

- Panel RE models use GLS which calculates a weighted average of the 'between' (differences between the companies' cost drivers) and 'within' (changes in the company's cost drivers over time) estimators. While OLS uses both estimators as well, it places a much greater weight on the between estimator than GLS which leads to different results.

- RE models require an assumption of time invariant inefficiency when decomposing the errors.

The calculation of the inefficiency estimation across all the models is an important consideration which we discuss further in Section 2.3.2. Depending on how the companies' inefficiency is calculated this may however be a moot point, i.e. in RE the inefficiency is calculated based on the error term as a secondary step, instead a ratio-based approach can be used which does not assume time invariant inefficiency (this is discussed further in Section 5).

2.3.2. Efficiency estimation

The different methods used to estimate the coefficients make different assumptions about how efficiency varies (or does not vary) over time, which we explained in more detail in our earlier report. They also use different methods to estimate coefficients. Here, the most robust models tended to be the GLS (RE) models, which assume that efficiency does not vary over the time covered, i.e. five years for water and seven years for sewerage. Although this may seem a rather bold assumption, it is supported by the SFA testing,[22] which allows for efficiency to vary in some systematic way (unlike OLS, which assumes that companies' efficiencies are not related over time but rather vary in a random manner).

In many cases the GLS (RE) models were preferred over OLS in terms of the signs, magnitudes and statistical significance of the parameter estimates. However, the assumption in the RE model of time invariant inefficiency, particularly when viewed over a seven year period, may appear rather restrictive. We therefore tested three additional, time varying panel models. The advantage over RE is that these models permit time varying inefficiency. The advantage over OLS in this respect is that the variation is structured over time, not time independent as in OLS.

The first two models are the BC92 and Cuesta (2000) models, which are both maximum likelihood stochastic frontier models. The first is commonly used in the literature, partly because it is easier to implement in standard software. The disadvantage of BC92 models is that they require all firms to have the same direction of efficiency change over time (that is, all firms see increasing or decreasing efficiency over time). The Cuesta (2000) model is more difficult to implement, and the Institute for Transport Studies (ITS), University of Leeds has developed LIMDEP (a statistical software package) code for this purpose. It has appealing properties in a

---

[22] This refers to the BC92 and Cuesta testing further below in this section.

regulatory context as it allows each firm to have its own time path for inefficiency, so some firms can be catching up to the frontier, whilst others may fall away.

The third model, CSS (1990), likewise permits firm specific time paths for inefficiency, but without the need to impose distributional assumptions (unlike the BC92 and Cuesta). One disadvantage of the CSS model is that it does not allow us to test the statistical significance of the time variation in inefficiency.

In general we found that the BC92 and Cuesta 2000 models were not robust. In many cases the models did not converge.[23] Where the BC92 models did converge, they tended to show that inefficiency was not varying over time. Finally, with both the BC92 and Cuesta models that did converge, there was some ambiguity concerning the estimation of the standard errors. This led us to conclude that these models should not be included in our suite of models (though we would suggest keeping them as possible approaches for PR19).

We also tested 'true random effect' models, which attempt to disentangle unobserved heterogeneity between companies and inefficiency by assuming that the unobserved heterogeneity is constant over time, while inefficiency is allowed to vary. However, as noted previously, this model can have difficulties distinguishing between persistent inefficiency and time invariant heterogeneity. We did not find these models to be viable as they yielded errors.

As a result, the final selection includes models using GLS (RE) and COLS respectively.

Box 2.2: Small sample performance of GLS (RE) and COLS

While GLS (RE) and OLS are similar approaches, as discussed above, they place different weight on the within estimator. There are numerous discussions around the merits of each of the approaches, but one area that can be an issue in a regulatory context is small samples. We discuss this further below.

While in small samples there is uncertainty about the performance of GLS (RE) estimators the academic literature indicates that GLS (RE) is no worse than FE and OLS.[24] In fact, GLS (RE) has been shown to outperform OLS and FE estimators in small samples (even in the presence of correlation between firm effects and regressors) due to its superior efficiency, i.e. preciseness of parameter estimates.[25] The benefit of having more precise coefficient estimates with GLS (RE) therefore may well outweigh the cost of having some correlation between regressors and firm effects (part of the residual). Any problems such as correlation between regressors and company effects would cause bias in OLS as well. Our extensive testing has suggested that the non-correlation assumption is reasonable. Furthermore, there are studies showing that GLS (RE) outperforms FE and OLS in small samples.

However, academic literature has shown in some cases that the superior efficiency becomes less favourable in samples where N-K<5, where N is the number of observation (in this case the number of companies as the variables have small within variation) and K is the number of

---

[23] Convergence in this case means that one of the criteria for exiting the iterative process of calculation within the statistical software were not met and the software could thus not generate model coefficients.

[24] See for example Taylor, W.E., *Small Sample Considerations in Estimation from Panel Data, Journal of Econometrics 13*, 2008, pages 203-223.

[25] See, for example, ibid; and Baltagi, B. H., *Econometric Analysis of Panel Data*, 2005.

variables, excluding translog terms. Therefore, in cases where N-K<5, the GLS (RE) estimators may not perform as well as expected.

Additionally, the way we understand Ofwat intends to use the models mitigates concerns about unobserved heterogeneity, 'within' variation, or correlation between drivers influencing the benchmarks. The calculation of average and/or upper quartile efficiencies in effect controls for the difficulty in distinguishing between unobserved heterogeneity and inefficiency (and noise in the case of OLS) by not using the frontier.

In practice, although GLS (RE) and OLS use different methods to estimate coefficients, their parameter estimates generally converge in our final set of models. Where they do not, the OLS estimates are within the confidence interval of the GLS (RE) estimates.

2.4. Panel length

The August submissions allowed us to extend our datasets for both water and sewerage by two years, thus allowing for a nine-year panel for water and an eleven-year panel for sewerage. However, Ofwat advised us that in the first two years of the sewerage dataset the costs were unusual because of a serious outbreak of foot and mouth in the preceding year. This meant that the costs and driver information during these two years was not consistent with the rest of the dataset because of the additional cost of disposing of the sludge or storing it for a longer period. Therefore, we reduced the length of the panel set to exclude the first two years in order to avoid this data consistency issue.

Because of the constraints of RE we were reluctant to fully rely on the longer panel as it would mean that companies' relative efficiencies would stay constant over seven years for water and nine years for sewerage. We therefore tested shorter panel lengths – five years for water and seven years for sewerage. However, as we discuss further in Section 5, the constant efficiency is not an issue when using an alternative method to estimate frontier or upper quartile efficiency challenges. In general, the long panel estimates were very similar to the short panel estimates. Where the model parameters were dissimilar, the long-panel estimates were within the short-panel confidence intervals.

We considered that the five-year panels for water were preferable given that there are 18 companies. However, as there are fewer sewerage companies (10 companies), we chose a seven year panel to allow for additional observations.

2.5. Smoothed versus unsmoothed capex

Capex in network companies is generally 'lumpy' over time, this is either due to the need to replace existing assets as and when needed or because expansion of a network is on a stepped basis rather than continuously. This means that capex does not generally move 'smoothly' in line with the cost drivers which causes difficulties with the modelling estimation. We believe that a partial solution to the problem is to use the smoothed capex, which would be interpreted as

annual capex on average over a given period.[26] We note that Ofgem used a smoothed capex approach for RIIO-GD1.

The lumpiness of capex for water and sewerage is illustrated at the industry level in Figures 2.1 and 2.2 below. These figures show that unsmoothed capex is lumpy and could possibly result in less robust results (and we note that at the company level capex is even lumpier). The figures also show capex smoothed over a five-year period. Given the length of the dataset available to us, we considered that smoothing over five years (which is also consistent with the price control length) was appropriate.

Figure 2.1: Water capex profile (£m real)
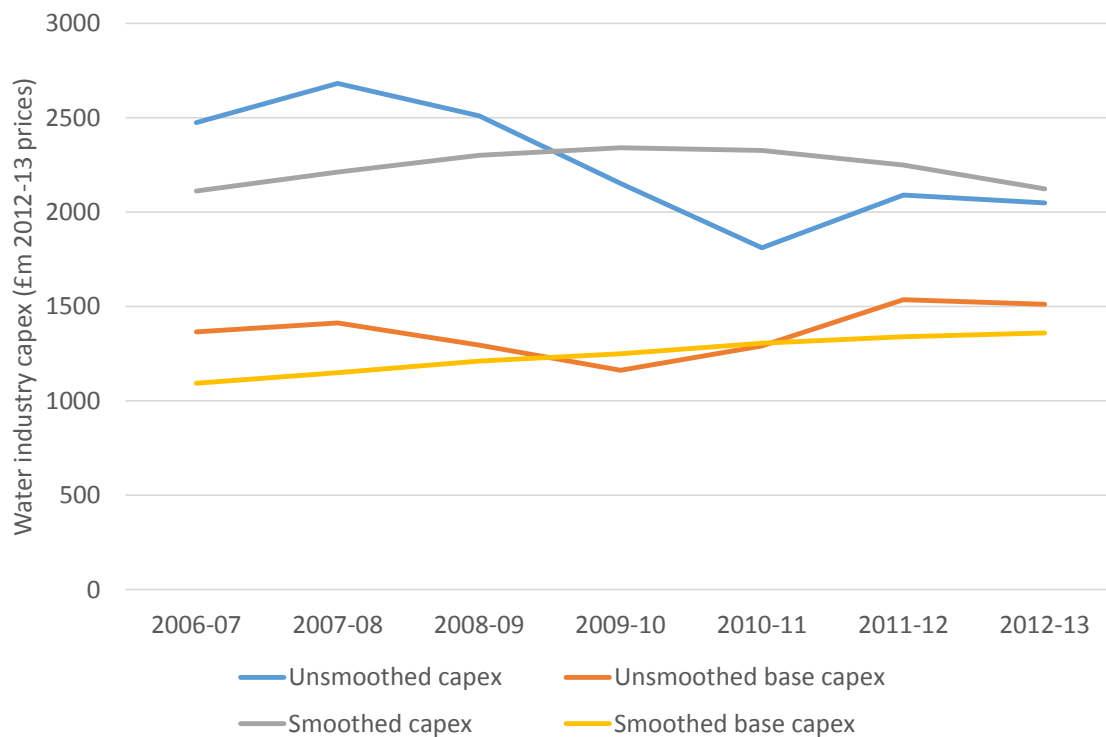
In sewerage, the average effect of smoothing base capex is even more pronounced – see Figure 2.2 overleaf.

---

[26] We note that there is regulatory precedence for using smoothed capex, for example Ofgem used seven-year smoothed capex for RIIO-GD1.

Figure 2.2: Sewerage base capex profile (£m real)
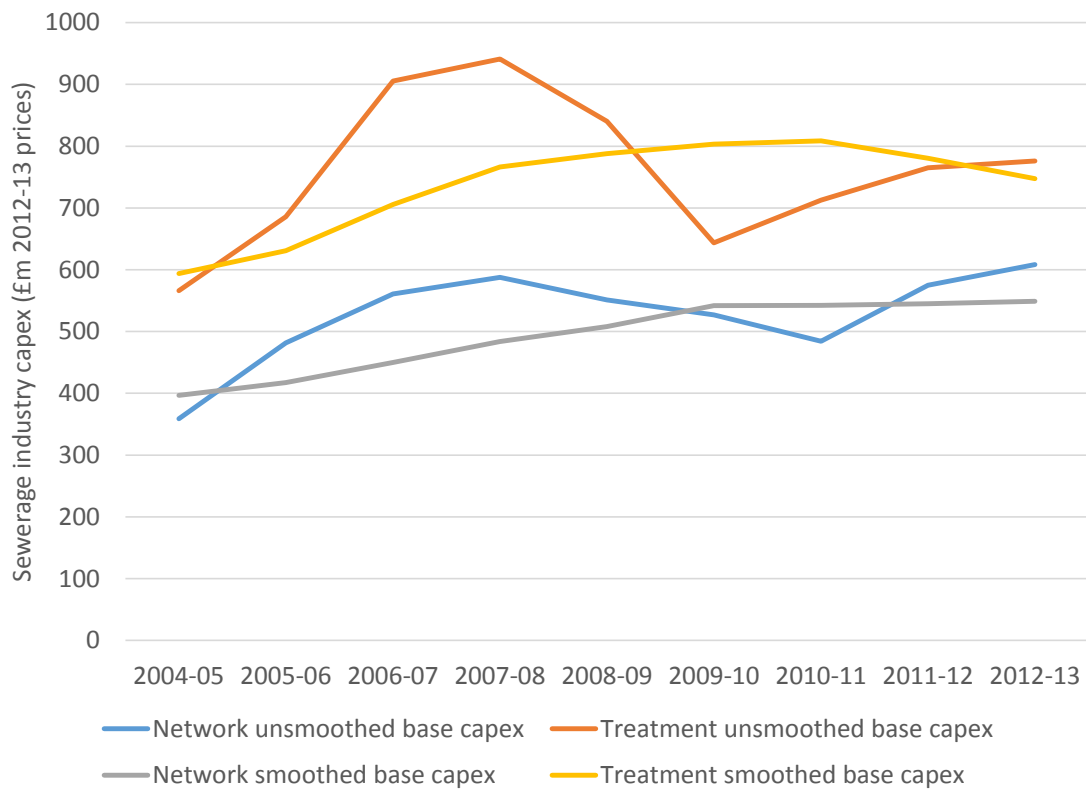
We tested the use of the unsmoothed capex measure as the dependent variable and found these models to perform less well than their smoothed capex counterparts. We used smoothed capex in all the models presented in this report.

3.   MODEL SELECTION CRITERIA

We developed multiple models at different levels of the water and sewerage value chains. We set out the initial viability testing of these models in our earlier report. As the model development set out in the earlier report dealt only with the specific question of whether totex or total cost models were viable we did not focus on a relative assessment of the different models. This meant that we had a range of models which varied by functional form, estimation method, variables included and transformations. In order to assess these models five standard criteria were used:

- theoretical correctness;

- statistical performance;

- practical implementation issues;

- robustness testing; and

- regulatory best practice.

Figure 3.1 briefly introduces our general logic in applying the model selection criteria. The following sub-sections discusses these criteria in more detail. While we have tried to keep the criteria as objective as practicable, given the nature of cost assessment modelling some element of subjectivity is required.

We also considered that there is a trade-off between the models, e.g. one model may have a more theoretically correct cost function while another may be more parsimonious and have more intuitively appealing coefficients. This may result in us recommending more than one model for use in setting the cost benchmarks and/ or baseline. The flowchart below (Figure 3.1) does not include practical implementation and regulatory best practice as, at this stage, we consider all our models to be relatively easy to implement and in line with regulatory best practice. However, we discuss these two criteria later.

Note, as set out in Section 2 of the CEPA *Cost Assessment Report*, the initial development of the models was undertaken with due consideration to Ofwat's Future Price Limits principles. Given that the models assessed in this report build on those initial models, we believe that each of the models assessed in this report are consistent with these principles.

Figure 3.1: Model Selection Process
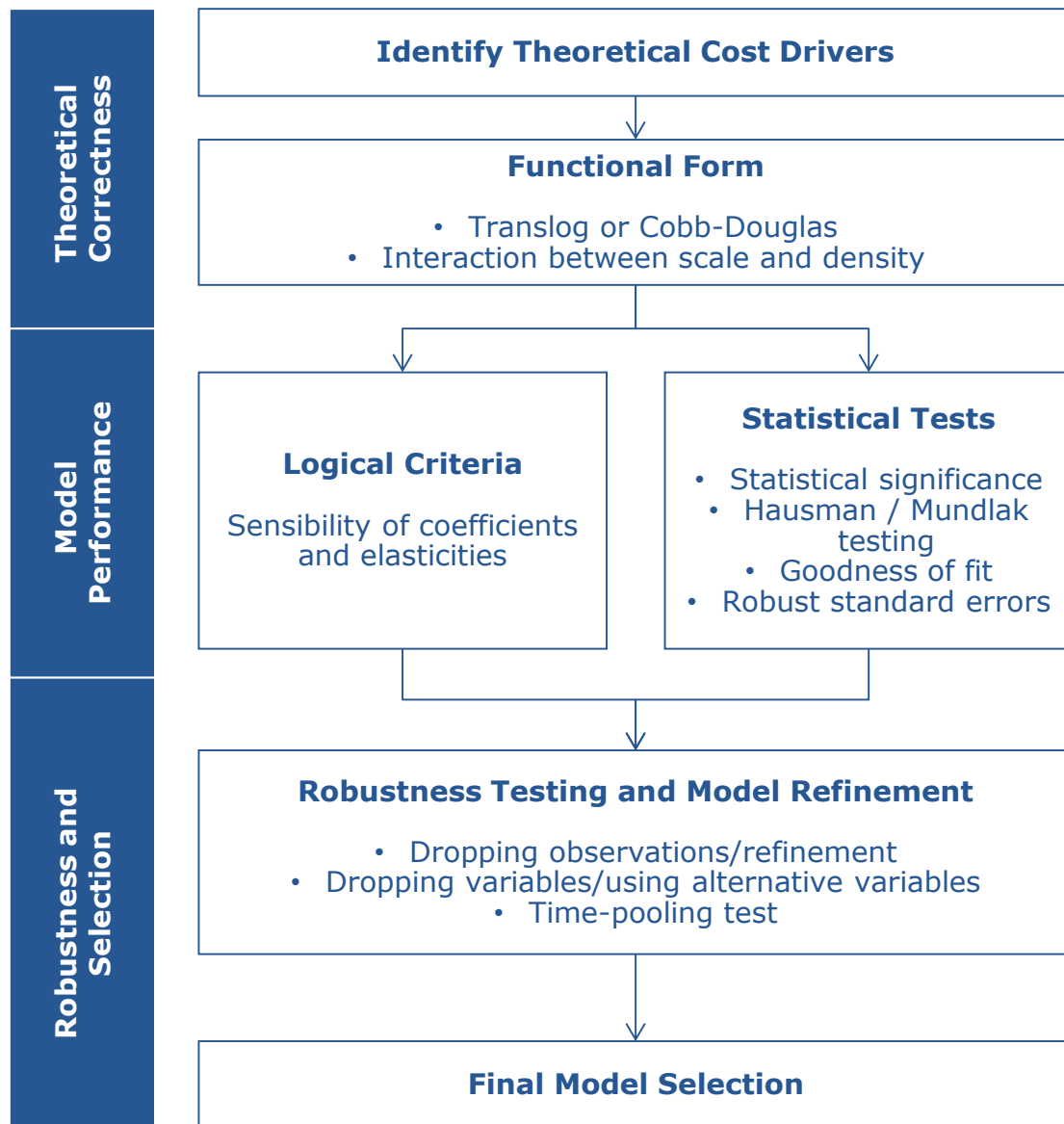
3.1. Theoretical correctness

3.1.1. Cost drivers

Theoretical correctness underlies all the modelling we have undertaken. In discussion with Ofwat,[27] we developed the models to reflect how companies' costs are driven. Therefore, theoretical correctness of the functional form (cost function) should ensure that the models reflect the underlying characteristics of the industry. However, it is important to bear in mind that models are always, to some extent, an abstraction from reality. The model estimation software provides statistical evidence as to whether the models fit the theoretical expectations. The main items considered in terms of theoretical correctness are CD versus translog and the efficiency assumptions.

---

[27] At the beginning of the project discussion also took place with UKWIR.

3.1.2. Functional form

Adopting a translog model (which allows for varying economies of scale across companies) allows for the changing nature of the economies of scale for the vertically integrated water and sewerage companies. As discussed earlier in Section 2.2, this theoretical assumption is consistent with earlier studies of the economies of scale in the industry. Translog models are, however, less transparent (we discuss the transparency issue in Section 3.4 'practical implementation issues' criteria) than other model forms.

CD linear models are easier to replicate, but suffer from the imposition of a single degree of economies of scale being assumed across the industry, i.e. all companies are assumed to face one of increasing, constant or decreasing returns to scale.

3.1.3. Time varying inefficiency

We also looked at whether a time-varying or a time-invariant efficiency is theoretically more suitable for the length of panel modelled. For longer periods, we would prefer to have time-varying efficiency models (COLS or SFA) as constant efficiency over a longer period of time could be a strong assumption (under RE). We note that this is only a concern if the model residuals are used to make efficiency adjustments.

Functional form cannot be considered independently from statistical performance of the variables in the models, which is discussed in the next criterion.

3.2. Statistical performance

3.2.1. Variables

The theoretical correctness should ensure that the variables included in the models can be justified as driving or affecting the level of costs and that they reflect the underlying characteristics of the industry. We reduced the range of variables included in the models by considering the following factors:

- **Statistical significance** – is the variable statistically significant? (to be weighed against the other factors below).

- **Sector significance** – is the variable one that *a priori* is expected to be an important explanatory variable?

- **Appropriateness of the result** – is the sign and impact of the variable what would *a priori* be expected?

With respect to the last criterion, considering the robustness of the explanation for any variable included was important. The latter two criteria are particularly important as focusing only on the statistical significance of variables may result in a mis-specified model due to multicollinearity, measurement error in the regressor, etc.

An important aspect affecting the statistical significance of the variables is the correlation between the explanatory variables. The higher the correlation between variables the less reliable the coefficients for these variables will be, and therefore they will also be less significant.

However, the overall predictive power of the model will be unaffected. We can chose between a parsimonious specification, which has the advantage of fewer variables that are more precisely estimated, and a fuller specification, which guards against omitted variable bias and unobserved heterogeneity, but results in coefficients being imprecisely estimated. If the focus is on efficiency measures (derived from the residuals between the estimated and the observed values), the latter may be preferable as it would take into account the full range of factors that affect costs and thus reduce the size of the residuals. On the other hand, this then may impede efforts to judge whether the shape of the frontier (determined by the parameter estimates) is plausible. We provide more detail on these matters in Section 3.3.1.

Furthermore, careful judgement must be exercised when considering the implications of leaving in a variable with an unexpected coefficient. We encountered a few model specifications particularly in sewerage, in which a few variables fell into this category. In general, we would be less concerned about a variable with an unexpected sign/size that is not statistically significant. However, we still had concerns about using the specification where a coefficient had a large unexpected value, even if it were not statistically significantly from zero, given the implications for predicting future expenditure.

In all the models we have taken the log of the explanatory variables (except for the dummy variables). Log-linear models reduce the risk of heteroskedasticity and allow for easier interpretation of the coefficients. The coefficients on the variables reflect cost elasticities, in other words if the coefficient on an explanatory variable is 1.0 then a 1% increase in the explanatory variable will lead to a 1% increase in the costs.[28] Log-linear models are the most common approach in academic and regulatory literature.

In Tables 3.1 and 3.2 below we set out our expectations for plausible ranges of the coefficients on explanatory variables for the water and sewerage models respectively (a more detailed description of the specification of variables is provided in Annex 1). The expectations are based on our in-team knowledge combined with input from engineers at Ofwat, initial UKWIR meetings with the industry cost assessment steering group and review of the academic evidence.[29] We set out these expectations on the basis of ignoring the effects of all other variables. We note that the ranges below may not apply in models with high multicollinearity between variables.

In translog models, the expectation of the magnitude of translog variables (i.e. squared and cross-terms) are less clear than coefficients on first order terms. There are a few reasons for this. First of all, when estimating at the industry sample mean, the squared and cross-terms cancel out such that elasticities at the sample mean are given by the first order term only. When examining elasticities away from the sample mean, these terms inform us of the curvature of the cost function. Therefore, although one may be able to have expectations on the magnitude of cost elasticities and whether these elasticities should be increasing or decreasing with a relevant variable, the speed at which the cost elasticities are changing (controlled by higher order terms) is not clear. Lastly, we note that in the past Ofwat has not used such translog variables in cost assessment, and thus it is harder to appeal to historical precedent to formulate expectations of

---

[28] Because we normalise all the translog variables to the sample mean the coefficient on the first order can be interpreted as the elasticity at the sample mean. We note that when the models are used to forecast expenditure, we use the coefficients that have not been normalised. This does not affect the predictive power of the model.

[29] For example see Stone and Webster 2004a and Saal et al 2011.

higher order terms in UK water and sewerage industries. Nonetheless, we did look at cost elasticities associated with these variables (away from the sample mean) but we refrain from including any expectations on magnitude or sign in the following table.

Table 3.1: Range of explanatory variables in water models

Type	Variable	Cost elasticity expectation
Core	Length of mains	These scale variables should be the main drivers of costs. Across these variables we would expect a value of above 0.7 and lower than 1.1.[30] A value above 1.0 could indicate diseconomies of scale/ density. In the models using a translog form, interpretations of the normalised coefficients are at the sample mean.
	Property density	
	Usage	
	Time trend	The time trend captures a combination of real price effects (RPE), changes in efficiency and changes in quality not explained by other explanatory variables. We would expect the coefficient to be relatively low, between -0.05 (~-5% per annum) and 0.05 (~5% per annum), as it is only picking up input price inflation above RPI.[31]
Input prices	Average regional wage	As labour costs make up a relatively high proportion of totex, we would expect the regional wage coefficient to be relatively high and positive, circa 0.6-0.7, but below 1.0. i.e., if wages were 1% higher in a company's region then we would expect overall costs to be higher but not by more than 1%.
	Regional BCIS index	The BCIS index effectively acts as a relative (regional) construction price indicator. We would expect the coefficient to follow the same logic for regional wages but to influence the remaining proportion of totex (that is not labour-related or determined at the national level), i.e., <0.4. This variable should not capture changes over time.
Network characteristics	Population density (occupancy)	As with the core scale variables, we would expect a coefficient of around 0.7 to 1.1.
	Proportion of metered properties	We would expect a relatively small negative coefficient, between -0.1 and 0.0, as metered properties are expected to have lower water consumption than non-metered and hence lower costs. If usage is included in the model it is not clear what the effect will be as the cost difference effect could be picked up in either or both variables. We have excluded this variable in the further model refinement because of the uncertainty of its effect on costs.
	Proportion of usage by metered household properties	We would expect a coefficient of around 0.4 to 0.9 (depending on the proportion of metered properties). (If usage is included in the model it is not clear what the effect will be as the cost difference effect could be picked up in either or both variables.)

---

[30] Competition Commission (2000), *Mid Kent Water plc: A Report on the References under Section 12 and 14 of the Water Industry Act 1991*, P 267, Professor Stewart, Ofwat's then academic advisor, estimated a cost elasticity of scale of 0.96.

[31] As this is a dummy variable, the coefficient needs to be adjusted using the formula exp(X)-1 to establish the percentage change in costs.

Type	Variable	Cost elasticity expectation
	Proportion of usage by metered non-household properties	We would expect a coefficient of around 0.4 to 0.9 (depending on the proportion of non-metered properties). (If usage is included in the model it is not clear what the effect will be as the cost difference effect could be picked up in either or both variables.)
Treatment and sources characteristics	Sources (number of)	We would expect a low positive number as taking water from more sources drives up costs.
	Pumping head (x distribution input)	This is used as an energy proxy. As energy is a significant driver of costs we would expect this to be relatively high, say 0.4 to 0.6.
	Proportion of water input from river abstractions	We would expect a low positive figure as water from abstractions is expected to lead to higher costs than water from boreholes (our excluded variable). However, this is not always clear because of bankside storage limitations.
	Proportion of water input from reservoirs	We would expect a low positive figure as water from reservoirs is expected to lead to higher costs than water from boreholes (our excluded variable).
Activity	Proportion of new meters	We would expect a low positive number as the installation of new meters should drive up capital costs.
	Proportion of new mains	We would expect a low positive number as the installation of new mains could drive up costs.
	Proportion of mains relined or renewed	We would expect a low positive number as the renewal/relining of new mains could drive up costs.
Quality	Properties below reference pressure level	We would expect a low negative coefficient as the lower the proportion of properties with inadequate water pressure the higher the capex costs would have been to reach that improvement in quality.
	Leakage	We would expect a low negative number as greater costs may be required to achieve a lower leakage level should leakage behave as a quality variable.
	Properties affected by unplanned interruptions > 3 hrs	We would expect a low negative number as greater costs may be required to achieve a lower level of properties affected by unplanned interruptions should this variable behave as a quality measure.
	Properties affected by planned interruptions > 3 hrs	We would expect a low negative number as greater costs may be required to achieve a lower level of properties affected by planned interruptions should this variable behave as a quality measure. This is an ambiguous driver as planned interruptions could also be a sign of quality improvement or scheduled maintenance.

Table 3.2: Range of explanatory variables in sewerage models

Type	Variable	Cost elasticity expectation
Core	Length of sewers	These scale variables should be the main drivers of costs. Across these variables we would expect a value of above 0.7 and lower than 1.1. A value above 1.0 could indicate diseconomies of scale. In the models using a translog form, interpretations of the normalised coefficients are at the central mean.
	Usage	
	Property density	We expect this to be a main cost driver. However, the sign of the density coefficient is expected to vary between network and treatment/ sludge models. In network models, we expect it to carry a positive coefficient due to increased costs associated with operating in urbanised areas. In treatment/ sludge models we expect a negative coefficient due to the ability to have larger, more efficient treatment plants serving densely populated areas. For these reasons, the expected sign of the density coefficient in combined models (capturing both network and treatment & sludge) is ambiguous.
	Time trend	The time trend captures a combination of real price effects (RPEs), changes in efficiency and changes in quality not explained by other explanatory variables. We would expect the coefficient to be relatively low, <0.05, as it is only picking up input price inflation above RPI.[32]
Input prices	Average regional wage	As labour costs make up a relatively high proportion of totex, we would expect the regional wage coefficient to be relatively high and positive, circa 0.6-0.7, but below 1.0. i.e., if wages were 1% higher in a company's region then we would expect overall costs to be higher but not by more than 1%.
	Regional BCIS index	The BCIS index effectively acts as a relative (regionally) construction price indicator. We would expect the coefficient to follow the same logic for regional wages but to influence the remaining proportion of totex (that is not labour-related or determined at the national level), i.e., <0.4. This variable should not capture changes over time.
Network activity	Proportion of sewers replaced and renovated	We would expect a low positive number as the refurbishment of sewers should drive up costs.
Treatment	Load	This scale variable for sewage treatment should be the main driver of costs. We would expect a value of above 0.7 and lower than 1.1. A value above 1.0 could be taken to indicate diseconomies of scale.

---

[32] As this is a dummy variable, the coefficient needs to be adjusted using the formula $\exp(X)-1$ to establish the percentage change in costs.

Type	Variable	Cost elasticity expectation
	Sludge disposed	As a possible substitute for the load variable we would expect similar values i.e. a value of above 0.7 and lower than 1.1. A value above 1.0 could be taken to indicate diseconomies of scale. Could also be considered a core variable as highly correlated with length.
	Proportion of load in treatment works size bands 1-3	We expect a positive coefficient on this variable as works in bands 1-3 tend to be more expensive than band 6 (the omitted proportion) in terms of unit costs due to economies of scale.
	Proportion of load in treatment works size band 4	We expect a positive coefficient on this variable as works in band 4 tend to be more expensive than band 6 (the omitted proportion) in terms of unit costs due to economies of scale.
	Proportion of works load in treatment works size band 5	We expect a small positive coefficient on this works density variable (if higher size bands are omitted in the model) to take into account the diseconomies of scale of band 5 works relative to band 6.
	Proportion of works load in treatment works size band 6	We expect a small negative coefficient on this works density variable if included in a model as it would take into account the economies of scale of band 6 works compared to the lower omitted band(s).
	Proportion of load undergoing activated sludge treatment	We expect a positive coefficient as this treatment is considered the most expensive treatment type.
	Number of large works with the tight consent dummy	Based on prior Ofwat large works models, this variable should have a coefficient around 0.1 to indicate higher costs associated with tight consents on ammonia, BOD_5, and suspended solids.

3.2.2. Hausman test

We used the Hausman test to choose between GLS (RE) and FE models. The test, a standard econometric test for model specification, indicates whether a GLS (RE) functional form is similar to FE. Similarity between GLS (RE) and FE indicated by the Hausmann test suggests the assumption of non-correlation between company effects and regressors in GLS (RE) is reasonable (as FE will always be consistent even when the non-correlation assumption breaks down).

In some cases LIMDEP cannot invert the variance-covariance matrix.[33] The LIMDEP manual indicates that the best interpretation of this leads to a conclusion that favours the GLS (RE) estimator (this was also supplemented by additional testing described below).

We also applied an alternative method for computing the Hausman test, known as the Mundlak approach. This approach is more general in its testing of correlation between company effects and regressors. The results of the Mundlak test broadly supported our findings from the earlier Hausman tests, reaffirming the preference for GLS (RE) models over FE. Where there were discrepancies between the findings of the Hausman and Mudlak tests we carried out further testing to isolate correlated variables (i.e. the variables causing the discrepancy between the two testing methods). Once isolated, we assessed the impact of controlling for correlation via the Mundlak approach. We note that controlling for correlation using the Mundlak approach makes the interpretation of coefficients more cumbersome and less transparent. We also found the impact of controlling for correlated variables to be small in sewerage models and produce unreasonable results in the water models.

Therefore, the general support of both the Hausman test and Mundlak approach for GLS (RE), the small differences when controlling for correlation when there were discrepancies between testing methods, and considerations of additional issues (e.g. transparency and interpretation of coefficients) led us to conclude that GLS (RE) is the preferred estimation method for our models.

3.2.3. Goodness-of-fit

Ideally we would have liked to assess the 'goodness-of-fit' of the models. Unfortunately, in GLS models there is no robust statistical measure of goodness-of-fit - see Green (2008).[34] As the majority of the models run for the water industry are based on a generalised least squares (GLS) estimator, the R-squared is not applicable. Furthermore, the R-squared tends to be high in log-linear models in general, which adds another layer of uncertainty to this statistic.

An alternative statistical measure of the goodness of fit is the square of the correlation between the observed and the predicted values of the models.[35] We note that this measure yields relatively high statistics and small differences in the statistics should not be used as indicating that a model is more robust. For example, a model with a 0.98 statistic should not be considered more robust

---

[33] This means that the differences of the two matrices is not positive and the Hausman statistic can thus not be generated. Greene provides more detail on this in the Limdep manual.

[34] Greene, W. H., *Econometric Analysis*, Sixth Edition, Pearson Prentice Hall, 2008, page 156.

[35] We have consulted William Greene on the most appropriate goodness of fit measure for GLS models.

than a model with a 0.97 statistic. We have also relied on the stability of the scores to robustness testing. While we have provided standard R-squared statistics for GLS (RE) models in Annex 4 and Annex 5 we warn against their use to avoid misinterpretation.

3.2.4. Robust standard errors

Robust standard errors refer to alternative ways of computing standard errors that try to take into account more complex structures within the data. In regular OLS estimation, variances of error terms are assumed to be a constant. However, it may be desirable to impose a covariance structure upon the error terms to take account more specifically for certain effects. White's robust standard errors take into account heteroskedasticity; that is different variances across different companies. Calculating robust standard errors has no impact on the parameter estimates themselves, only on the estimated standard errors and significance of parameter estimates.

White's standard errors were used consistently in OLS estimation as the assumption of a constant variance is unreasonable. White's errors were also tested in place of the standard errors calculated via GLS for the random effects models. It was found that these robust standard errors were similar to the GLS standard errors in terms of precision in most cases and would have led to equivalent choices of model selection. Greene also warns against using robust standard errors for GLS as their interpretation is not necessarily straight forward.

3.3. Robustness testing

We carried out several robustness tests, which included removing variables, dropping observations, statistical testing, changes in predictions, and rank correlations with other CEPA models.

3.3.1. Refinement

To get to the selected set of models, we refined them down from the full model specification by removing variables one at a time. We started by removing the non-core variables with the highest p-value (lowest level of significance) until we got to a stable model. This robustness check resulted in the refined models. We also checked the impact of dropping variables on coefficient estimates. We tried to include as much of the value chain as possible, which led to leaving in some variables even if they were not statistically significant.

Further refinement was necessary when, despite being statistically significant, the magnitude and/or sign of a variable was highly different from our *a priori* expectations (for example BCIS, discussed below). In those cases, besides looking at the coefficients, we also assessed the rank correlations and compared predictions of models covering the same cost area.

We found that the inclusion of two variables which we considered important cost drivers during the earlier phases of this project had unexpected results. These variables were:

- BCIS – in both water and sewerage; and
- Usage – sewerage only.

We discuss our findings with respect to these variables in Text Box 3.1 below.

Text Box 3.1: BCIS and usage

BCIS

All models explicitly take into account regional price differences based on the average regional wage variable and/or the BCIS variable, included on the right-hand side of the equations. This differs from Ofwat's approach in PR09 in which it made an ex-ante adjustment to modelled opex using regional wages and to modelled capex using BCIS.

We found, unsurprisingly, that the regional wage variable and the BCIS are highly correlated and when both are included in the modelling it resulted in odd coefficients (e.g., large and/or negative) and did not improve the predictive power of the models. We found that dropping the BCIS variable brought the coefficients on the other variables more in line with our expectations. We therefore dropped it in a number of models and relied on the average regional wage variable.

Usage

A similar case was made for the usage variable in the sewerage network model. Both OLS and GLS (RE) returned negative coefficients, statistically significant in the case of OLS. This implies that higher levels of usage decrease costs, opposite to what is expected.

The result was robust to model refinement as well; dropping BCIS increased the magnitude of this effect. Excluding usage from network had little effect on the models' predictive power, and brought other point estimates more in line with expectations. For these reasons, usage was also dropped from the network model.

Although both BCIS and usage are theoretically important *a priori*, it is clear from our estimations that there were significant problems with the variables. It is important to note that these variables are imperfect proxies and they may in fact be picking up undesirable effects of other included (or excluded) variables. In the case of BCIS, the data is not comparable year on year and only serves to proxy regional differences in construction prices within the year. It is also highly correlated with wages, the other regional price variable. In the case of usage, the variable tested is defined as load entering system/property. Since load is a measure that captures both the strength of the effluent and its volume, it is impossible to separate the effect attributed only to volume, which is the driver that applies to network activities. Usage performs better in the full wholesale base model, which is less susceptible to outlier observations and includes treatment costs, driven by the strength as well as the volume of sewage. While recognising the importance of scale and regional price variables in the models because of the above reasoning it seems reasonable to drop both the BCIS and usage variables.

In terms of rank correlations, we checked if the efficiency rankings of a model were consistent with those of the other models that covered the same part of the value chain or have the same type of expenditure (e.g., base expenditure, or base plus enhancements). This meant comparing:

- totex model results;

- sewerage network model results;

- treatment & sludge model results; and

- opex plus base capex model results separately.

Rankings and scores that were consistent with other models supported the robustness of our analysis for that particular part of the value chain/expenditure level. However, we note that different estimation methods may make different assumptions about efficiency, which may lead to diverging results. Consequently it was important that these results were discussed with Ofwat and robust judgements formed based on sector knowledge as well as modelling tests and these discussions were an important part of the development and testing process.

3.3.2. Dropping observations

We tested the sensitivity of the models' outputs by dropping observations. This tested the stability of our coefficients, efficiency scores and for the presence of outliers. We used rank correlations and predictions to compare our models. We preferred models that are less sensitive to outlier observations.

3.3.3. Pooling test

A structural break occurs if the effect of a cost driver changes from one period to the next. We therefore investigated two different scenarios where we thought a structural break was most likely to occur: the onset of the financial crisis and the beginning of the current price control (AMP5).

It is important to note that any variable may be tested for a structural break whether it is justified or not. Therefore, we limited our analysis to variables we thought could display a break from a theoretical/logical standpoint. We chose to investigate the BCIS index, regional wages, usage (sewerage only), and number of sources for water (only tested against AMP5). The first two were directly impacted by the financial crisis through pressure on input prices as demand slowed. The latter two are related to differences in regulatory reporting requirements between AMP4 and AMP5.

We concluded from our testing that there was no evidence of AMP5 affecting the parameters associated with our chosen cost drivers. In general, the onset of the financial crisis did not result in significant sensitivity of the coefficients of our chosen variables (i.e. the interaction term was not statistically significant). There was, however, evidence that the onset of the financial crisis did change the way in which regional wages drove costs in one of our models. Where this was the case, the effect had a negligible impact on forecasts, parameter estimates, and efficiency scores. Furthermore, we note that due to choosing a shorter panel length aimed at alleviating concerns of constant efficiency assumptions in the RE model, the 'pre-crisis' coefficients in the water models were based on a single year of data. This reduces the robustness of the 'pre-crisis' result. It is for this reason and the negligible impact on results that we concluded that the models were not sensitive to time-pooling.

3.4. Practical implementation issues

We considered that any proposed cost models should be transparent, replicable and stable. This includes ensuring that the models are not too complex (although this potentially involves a trade-off with accuracy and theoretical correctness), that the implications of the results are clear and the results of the models are objectively reproducible where applicable. We believe all the models

we included in the final round of testing are not unduly complex and can be implemented using standard econometric methods and software.

3.5.　Regulatory best practice

When developing new cost assessment models it is appropriate to review how other regulatory agencies carry out similar analyses. While we considered that checking the modelling methodology with that used by other regulators is useful, a different approach may not necessarily be a cause for concern as the data availability and context in which the analysis is undertaken may vary.

We believe the modelling we carried out offers benefits over Ofwat's previous cost modelling and is more in line with regulatory practice seen at other regulators, e.g. Ofgem and ORR. In particular, the approach utilises panel data, which is advantageous for a number of reasons (inter alia, it increases the sample size, enables variation in efficiency and technical change over time to be studied, and enables efficiency estimates to be derived without recourse to distributional assumptions).[36] We also note that the use of a panel data set is in line with the CC recommendations in the Bristol Water case.[37] ORR and Ofgem have both developed panel data models for use in their efficiency determinations, for example, Ofgem's RIIO-GD1 and RIIO-ED1. The approach is also in line with that of other regulators in seeking to benchmark total costs (or totex), or at least substantial parts of total costs together, rather than separately. Whilst this could potentially have some disadvantages compared to the more disaggregated approach taken by Ofwat in previous price reviews, in that more tailored models could be developed for different cost categories, it has major advantages in terms of addressing potential incentives for capital bias and ensuring that substitution between different categories of expenditure is taken into account. We note that Ofgem used (and is using) totex benchmarking, in combination with bottom-up benchmarking, for RIIO-GD1 and RIIO-ED1, and in PR08 ORR benchmarked maintenance and renewals together (although they did separate assessments for enhancements and operating costs).

Finally, we have used the same data (June Returns) as Ofwat has used for its previous cost modelling, plus the data submitted by companies in August 2013. With respect to our models we have tested a wider range of variables than covered in Ofwat's previous work, including quality measures, and our final models may be favourably compared with previous Ofwat models in terms of the number of variables included and the extent to which the coefficients accord with engineering understanding while also being statistically significant.

3.6.　Results coding

There is no singular method or metric for identifying suitable models mechanistically, rather a judgement is required in model selection. To facilitate this process, we have adopted an approach based on a 'traffic-light' system to indicate how well the model performs against a given criterion, i.e., a 'green light' corresponds to 'good', 'amber light' corresponds to 'acceptable but with a few issues', and a 'red light' means that the model is flawed.

---

[36] CEPA and Mott McDonald. *Cost assessment – use of panel and sub-company data.* May 2011.

[37] Competition Commission. *Bristol Water Plc Price Determination*. 2010.

In this sub-section we describe the method of assigning traffic lights to a short-list of models. The selection of traffic lights is based on the conclusions for each model summarised in the templates set out in Annex 4 for water and Annex 5 for sewerage. We note that we ran a much more exhaustive range of models than those presented in these annexes, but we pre-selected these as the most viable models.

As we mentioned earlier in the report, all the models presented here are in line with regulatory best practice and there are no obvious concerns about their practical implementation. We therefore only assigned traffic lights for the remaining three categories, i.e. theoretical correctness, statistical performance, and robustness checks. We considered whether the model meets a set of criteria for each category, listed by priority in the table below. The boundary between Amber and Green depends on whether the model satisfies the top criteria.

At this stage, we did not assign a red light to any model for theoretical correctness as the models had already been narrowed down to include a set of theoretical drivers following discussions with Ofwat, UKWIR and by implementing standard econometric approaches. The other categories – statistical performance and robustness testing – do allow for a red traffic light, in which case the model would no longer be considered a candidate. For the former, a red light indicates that several of the core parameter estimates are substantially outside the expectations in Tables 3.1 and 3.2 and are statistically significant. For robustness testing it means that either the efficiency scores resulting from the model or the prediction are implausible; or that there is significant evidence for having different coefficients in different time periods.

We considered that any model that received a red light (in any category) should not be used to set cost benchmarks/ baselines.

Table 3.3: Traffic light criteria in order of priority

	Theoretical correctness	Statistical performance	Robustness check
R	N/A	The core parameter estimates are substantially outside the expectations in Tables 3.1 and 3.2.	Overall range of efficiency scores and predictions is not plausible. Pooling tests suggest significant and material differences in coefficients for key variables in different time periods.
G **A**	1. Prefer translog over CD functional form, particularly for water where the models are not disaggregated by value chain and there is greater size variation between companies. Preference is based on theoretical reasoning and statistical significance tests of the translog terms. *Translog models given Green and CD given Amber, if translog is significant.* 2. Are all core theoretical drivers included? *If not, given Amber.*	1. Coefficient estimates largely in line with expectations (based on Tables 3.1 and 3.2) and elasticities relatively sensible. *If not, given Amber.* 2. How refined is the model? (Statistically significant parameter estimates while including as much of the value chain drivers as possible.) Is N-K >5 for RE?[38] *The most refined models given Green.* 3. Statistical results: goodness of fit/ statistical preference for GLS (RE) over FE. *If FE preferred, given Amber.*	1. Sensitivity to dropping observations/ variables. *If efficiency scores or predictions are sensitive, given Amber.* 2. Are model rankings outliers with respect to other CEPA models at same level of expenditure and value chain disaggregation (see Annex 4 for details)? *If so, given Amber.*

---

[38] Used as a rule of thumb rather than a hard and fast rule, as we recognise there is no definitive threshold for reduced reliability of GLS (RE) estimates.

4. MODEL SELECTION

4.1. Introduction

In this section we focus on the models we determined to be the most viable, namely using GLS (RE) or OLS only, and then assess these models against the criteria set out in the preceding section. We do this in turn for water and then sewerage.

4.2. Water

4.2.1. Short list of viable water models

We narrowed down our preferred range of viable water models to 10. Seven of these models are at the totex level, while three use opex plus base expenditure. We summarise all these models in templates in Annex 4. The templates provide the results from our testing, coefficients and confidence intervals. A brief description of these models and our assessment of them against our criteria is set out in Table 4.1 overleaf.

Table 4.1: Select water models assessed

Model reference	Description	Theoretical correctness	Statistical performance	Robustness check
Totex				
WM1*	Fully specified totex GLS (RE) (translog); includes all theoretical water drivers.	G	R	A
WM2*	Fully specified totex GLS (RE) (translog), but excluding regional BCIS.	G	R	A
WM3	A COLS version of WM2.	G	A	A
WM4	Refined totex GLS (RE) (CD); variables included are length of mains, property density, time trend, regional wage costs, population density, proportion of input from river abstractions, and from reservoirs.	A	A	R
WM5	Refined totex OLS (translog); variables included are length of mains, property density, time trend, regional wage costs, population density, proportion of input from river abstractions, and from reservoirs.	G	G	G
WM6	GLS (RE) version of WM5.	G	G	G
WM7	GLS (RE) version of WM5 with BCIS included.	G	R	G
Opex + base capex				
WM8	Refined opex plus base capex GLS (RE) (translog); variables included are length of mains, property density, and their corresponding translog terms, time trend, average regional wage, regional BCIS index, population density, leakage, planned interruptions, proportion of input from river abstractions, and from reservoirs.	G	R	G
WM9	OLS version of WM8, excluding BCIS.	G	A	G
WM10	GLS (RE) version of WM9.	G	G	G

* Note, while the GLS (RE) fully specified models ran in our statistical programme (LIMDEP) because of the number of explanatory variables exceeded the number of companies it was not clear how the between estimator was calculated. Consequently we considered that the models failed the 'Statistical performance' criteria.

4.2.2. Water models recommended for triangulation

After giving due consideration to each of the models in Table 4.1, and in discussion with Ofwat, we recommend using a range of specifications (i.e., full and refined, and totex and opex plus base capex). We found that the full and refined tended to give slightly different results, but given the trade-offs of a richer model (full) and parsimonious model (refined) discussed earlier, there was no overwhelming reason for preferring one over the other. While the totex model offers the benefit of not requiring unit cost models for the enhancement capex the opex plus base capex model appeared robust and offered an alternative view on the companies' efficiency. In a similar vein, other than the GLS (RE) models being slightly more robust than the COLS models in most cases there was no clear evidence why one should be preferred over the other. As the models provide different predictions we believe that using both estimation techniques is appropriate.

We recommend using the following five models, which are based on GLS (RE) and OLS versions of three basic model specifications:

- Full totex (**WM3**): As it included all the variables we considered to be theoretical drivers, this model is less likely to suffer from omitted variable bias than the refined models. The unexpected results for statistical significance and size/signs of the parameters may be due to multicollinearity, which would not pose issues for the overall predictive power of the model. The Amber in the robustness check category refers to the models' sensitivity to dropping variables, which we do not consider to be a drawback for a fully-specified model. As explained earlier models excluding BCIS are more appropriate given the correlation between this variable and regional wage.

- Refined totex (**WM5 and WM6**): The coefficients are generally as expected and the models have a high rank correlation, despite using different estimation methods. These models have advantages over the full model in that they are more parsimonious and the coefficients should be more precise.

- Refined base expenditure (**WM9 and WM10**): Although we prefer totex to avoid capex bias, these opex plus base models are sufficiently robust and in line with expectations to be used in triangulation, along with a unit cost estimate of enhancement. The amber in Model 9 reflects the unexpected coefficient on population density, which could be due to multicollinearity. We consider that the models can be used directly in triangulation or as a cross-check for the other totex models.

Comparing the efficiencies of the two refined water models, one can draw conclusions about the difference between base (WM10) and enhancement expenditure (included in WM6). In the base model, companies seem to be slightly closer to the average industry efficiency than in the totex model. This suggests that companies may differ more in the efficiency of their enhancement activities compared to base activities, though this could also be explained by greater variability in heterogeneity of enhancements. We can see this in Figure 4.1 below; it illustrates the range of efficiencies for the 10 companies that are closest to the industry average.
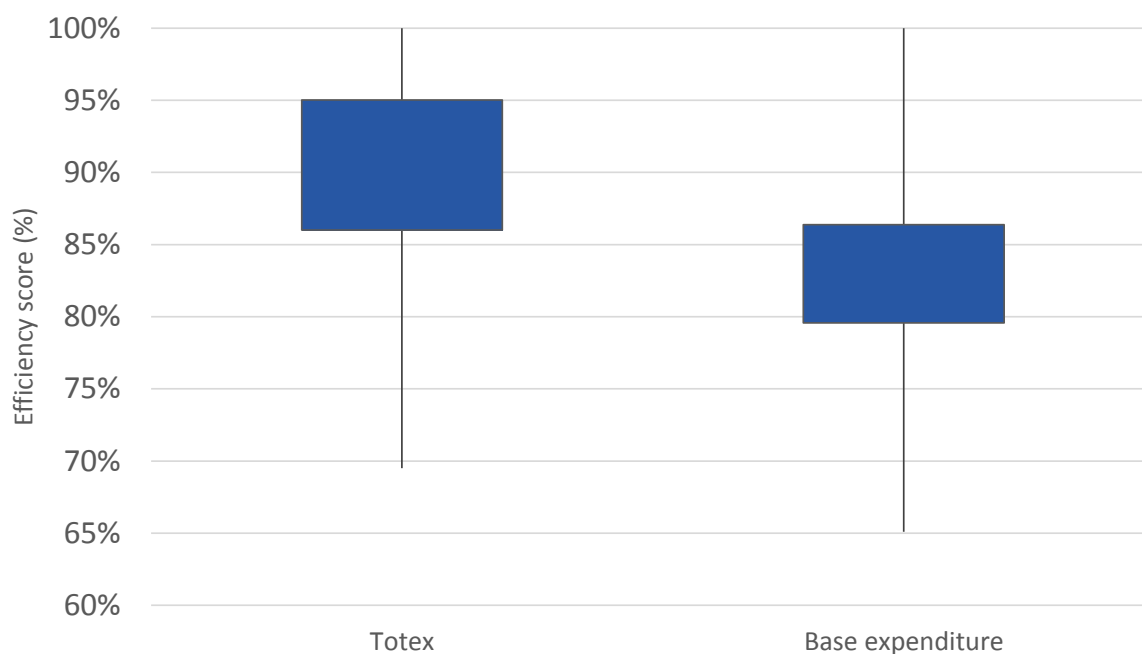
4.3. Sewerage

4.3.1. Short list of viable sewerage models

The models we tested in sewerage ranged from sewerage totex models to size-band sub-company models for sewage treatment opex only. As noted in the introduction to this paper, we dropped the sub-company models because, while viable, they failed to capture the linkages across the treatment activity achieved by a more comprehensive model.[39]

We narrowed our preferred range of models to 10. Two of these models were for network opex plus base capex, four for treatment and sludge opex plus base capex, and four for sewerage wholesale opex plus base capex. We did not identify any viable models which included enhancement capex. A brief description of these models and our assessment of them against our criteria is set out in Table 4.2 overleaf. We summarise all these models in templates in Annex 5.

---

[39] We considered this as a solution only when more encompassing models did not appear viable.

Table 4.2: Select sewerage models assessed

Model reference	Description	Theoretical correctness	Statistical performance	Robustness check
Network opex + base capex				
SM1	A refined translog GLS (RE) model that covers network base expenditure (opex and base capex); variables included are length of sewers, property density, and the corresponding translog terms, time trend and regional wages.	G	G	G
SM2	The OLS version of SM1.	G	R	G
Treatment & sludge opex + base capex				
SM3	Fully specified treatment & sludge translog GLS (RE).	A	R	R
SM4	Slightly refined treatment & sludge CD model (GLS [RE]); variables included are load treated, time trend, regional wages, proportion of load treated by activated sludge, proportion of load treated in size bands 1-3, sludge disposed.	A	G	R
SM5	A refined treatment & sludge GLS (RE) model that also uses a translog form; variables included are load treated, property density, and the corresponding translog terms, time trend and regional wages.	G	G	G
SM6	The OLS version of SM5.	G	G	G
Wholesale opex + base capex				
SM7	Fully specified translog GLS (RE) that covers both network and treatment & sludge.	G	A	A
SM8	The OLS version of SM7.	G	A	A
SM9	A refined version of SM7; variables included are load treated, property density, and the corresponding translog terms, time trend, regional wages and proportion of load treated in size bands 1-3.	G	G	A
SM10	A refined version of SM8 (this is also the OLS version of SM9).	G	G	A

4.3.2. Sewerage models recommended for triangulation

As with the water models after giving due consideration to each of the models in Table 4.2, and in discussion with Ofwat, we recommend using a range of specifications (i.e., network, treatment and sludge and sewerage wholesale). Aside from the network models, other than the GLS (RE) models being slightly more robust than the OLS models, in most cases there was no clear evidence why one should be preferred over the other. As the models provide different predictions we believe that using both estimation techniques is appropriate. For the network models, the OLS based model contained unexpected coefficients on the wage variable. This coefficient was highly negative and as such we had concerns about its interpretation and impact on the forecast predictions.

We recommend using five final models in sewerage. We note that none of these models cover enhancement, unlike water. The final cost benchmarks/ baseline estimates will need to be based on these models triangulated with the unit cost models to account for enhancement. The majority of the expenditure in sewerage is treatment and sludge related. The suite that we recommend is thus more treatment and sludge oriented in terms of explanatory variables. The final selection for triangulation covers the following models:

- Network (**SM1**): This is a refined network only model. It includes purely network related variables (e.g. length of sewers). The model uses GLS (RE). We believe it is a useful addition to the suite of final models along with the separate treatment models as it offers a bottom-up approach. We did not include an OLS model here as it included an unexpected coefficient on wages.

- Treatment and sludge (**SM5 and SM6**): These are two treatment and sludge only models, both of which are refined. These models include key treatment variables, some of which also relate to network to account for possible trade-offs in expenditure between the two business lines.[40] Full models did not add much to the predictive power in this part of the value chain (e.g. sludge disposed is highly correlated with load treated). As treatment comprises a significant portion of expenditure, we selected two models here, which provide a range of approaches (GLS [RE] and OLS). These models need to be combined with the network model (SM1) before they can be compared to the wholesale base sewerage models.

- Wholesale base sewerage (**SM9 and SM10**): These are models that cover the entire sewerage value chain (network and treatment and sludge). They cover the same range of drivers as the network and treatment and sludge models, but the range of variables are more treatment-oriented to account for the higher proportion of expenditure in treatment (therefore load is preferred to length as the key cost driver). These models are refined and did not appear to suffer from multicollinearity. Their predictive power is not very different from that of the full models. The key advantage of these models is combining network and treatment and sludge, which picks up any trade-offs between these two parts of the business. The only difference between these two models is the

---

[40] For example, there may be a trade-off between having a longer network with one large treatment plant or shorter networks with many small treatment plants (larger treatment plants are usually seen as more efficient).

estimation method, which leads to the low rank correlation between the two models, marked with amber in the robustness check category.

We believe this range of sewerage models accounts for several issues in sewerage: trade-offs between network and treatment and estimation method differences between GLS and OLS. However, we note that the trade-offs between these two areas are likely to be less 'dynamic' in nature, as the coefficients reflect the historical structure of the sewerage system, and if the models at the disaggregated level of expenditure contain the appropriate cost drivers the trade-off issue should be relatively minor.

In terms of average efficiency, the models demonstrate that most companies perform differently in network and treatment and sludge. The dispersion in treatment and sludge (SM5) is much higher than in network (SM1). In the combined wholesale base sewerage model (SM9), those differences diminish (in particular the spread between upper and lower quartile) because companies that were less efficient in one service often compensate by being more efficient in the other. We also note that in terms of the average efficiency level in the industry, the wholesale base sewerage model is more in line with the treatment model than with the network one as treatment accounts for the larger proportion of expenditure.

Figure 4.2: Sewerage efficiency ranges
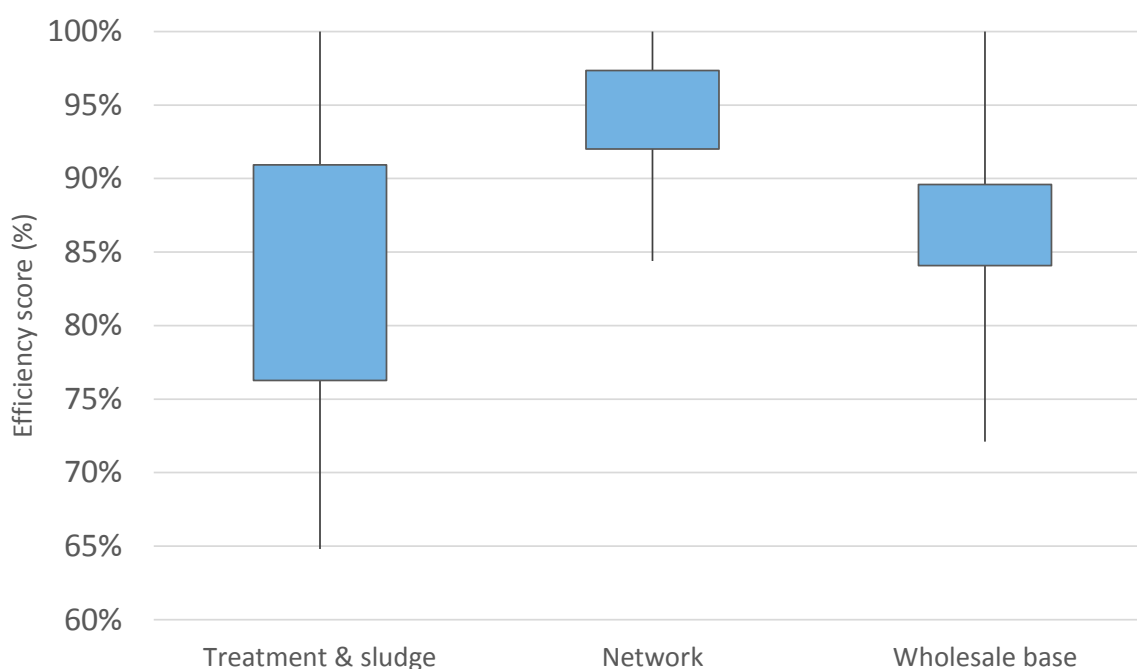
4.4. Other considerations

4.4.1. Time trend

The time trend variable in all the econometric models accounts for the frontier shift, RPEs and changes in quality not captured via the other variables in the model. A positive time trend indicates that the improvement in technology which would lead to savings had been outweighed by RPEs or increases in quality that the industry has paid for. A negative time trend indicates that gains in ongoing efficiency outweigh the other two factors put together. In previous price

controls Ofwat has applied RPEs net of ongoing efficiency of between 0.25 (for base opex) and 0.4% (for base capex).

In our preferred water models, time trends in totex are not statistically different from 0%, while at the base expenditure level they are around 1%. This could indicate a range of things, including that ongoing efficiency gains in enhancement have been greater than in maintenance and opex, or that expenditure related to improving quality is contained in maintenance and opex.

In sewerage, we only modelled base expenditure. We see a time trend of around 2% in both network and treatment and sludge. A possible explanation as we understand it, is that over AMP5 quality in sewerage has been improving and this would likely lead to higher costs in opex and base capex.

4.4.2. Economies of scale

Our modelling results show that there are varying returns to scale/density in both water and sewerage. This is allowed for by the translog specification, which was jointly significant in all models.

In water, elasticities with respect to length of mains (size) range between 0.9 and 1.1, suggesting economies of scale for some companies and diseconomies for others. The range is, however, tight with the average showing relatively constant returns to scale. In sewerage, all companies have elasticities with respect to size less than one, suggesting economies of scale.

It is also interesting that in terms of density, water and sewerage show different shapes of the elasticity curve. We find the extent of returns to density increasing in sewerage and decreasing in water. These results can be interpreted as having a more dense network facilitates treatment in large works in sewerage. In water, the density affect seems to be related to higher costs of maintenance work in urban areas.

5.   TRIANGULATION

We understand that Ofwat will use the econometric models to forecast the cost benchmarks for the risk-based review in PR14. Given that we were unable to narrow our preferred range of models to a single model for either water or sewerage, we recommend that the results from the preferred list of models be weighted together. We refer to this approach as 'triangulation' and we briefly discussed it in the CEPA *Cost Assessment Report*. We note that, where the models do not use totex the results from the unit cost models and any non-modelled costs must be added to achieve a view of the companies' totex.

The raw model estimates may also require an adjustment to avoid log-transformation bias.[41] We discuss this in more detail in Annex 7 and we recommend the use of either the 'alpha factor' or 'conditional mean' but we consider that the final choice of adjustment is up to Ofwat. We note that these adjustments should be applied before triangulating.

We also note that while Annex 4 and Annex 5 show the models' coefficients at the sample mean for comparison purposes in model selection, the non-normalised coefficients should be used in forecasting AMP6 expenditure. Non-normalised coefficients are the ones resulting from modelling that uses data in which the three translog variables have not been divided by the average of the sample. Annex 8 provides those coefficients that are to be readily used in Ofwat's feeder models and provides further explanation of how those are reconciled with the normalised coefficients.

5.1.   Triangulation options

There are a number of ways in which one could triangulate the models' predictions to yield a final cost benchmark or baseline value. We therefore focused on methods based on the following logical flows:

1. Triangulating based on estimation method to arrive at GLS (RE) water (sewerage) and COLS water (sewerage) estimates that are then combined.

2. Triangulating across disaggregated models to reach a single bottom-up 'totex' value and then combining this with a single value from top-down 'totex' models.

3. A combination of Option 1 and 2. Triangulate based on estimation method then bottom-up vs. top-down. This gives us bottom-up (top-down) GLS (RE) and COLS estimates that are then combined.

4. Similar to Option 2, we build 'bottom-up' and 'top-down' estimates first, but keep a distinction between refined and full models.

While in practice there is little difference between the results of the triangulation process, we considered additional criteria that led a single recommendation. These criteria are:

- The intermediate information each option offers - i.e. the usefulness or intuition of information contained in each step.

- Transparency.

---

[41] This is due to Jensen's inequality.

- Logical flow i.e., do the weights make intuitive sense.

- Ease of implementation/ replicability.

Following discussions with Ofwat we concluded that Option 4 best met the criteria set out above. We considered that the preservation of a bottom-up estimate provides useful information from a business plan perspective while being weighted with the encompassing view of a top-down totex model. Furthermore, it maintains a logical split between full and refined top-down models for water only companies. In addition, we believe that the implicit weights applied to each model in this triangulation method are intuitive and logical.[42] Option 4 is illustrated in Figure 5.1 and Figure 5.2 for water and sewerage respectively.

Figure 5.1: Water triangulation

---
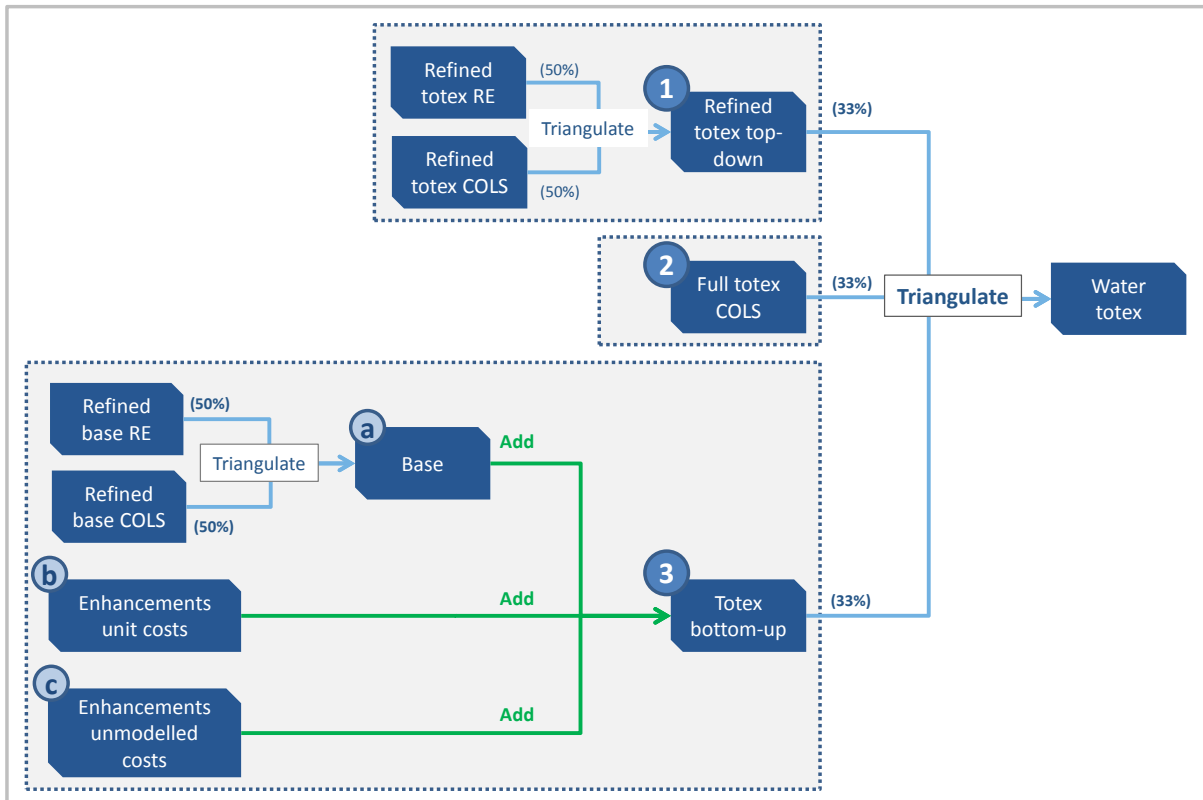
[42] Though the option to set explicit weights remains available, we considered that this would only be required if new information became available suggesting a preference between the aggregate/ disaggregated models.

Figure 5.2: Sewerage triangulation
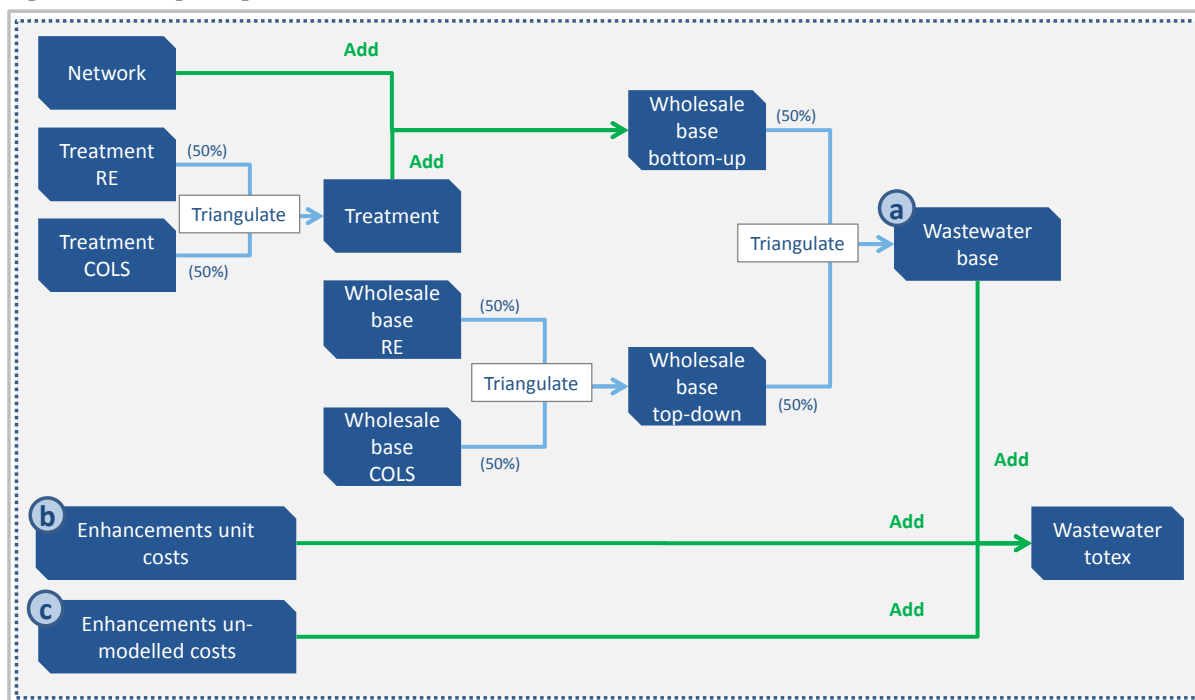
5.2. Efficiency adjustments

Model cost estimates are all calculated at the average industry efficiency, and there are several ways of making adjustments to these projections when setting efficiency targets. In essence, they are all ways of shifting the prediction line to the upper quartile (UQ), lower quartile (LQ), or frontier.[43] Here we give an example with the upper quartile but the same logic applies to the other adjustments.

5.2.1. Method: ratio- or residual-based

We consider two different methods of calculating the adjustment for upper quartile efficiency:

- based on the residuals from each model; or

- based on the ratio of actual expenditure to predicted expenditure.

We discuss this in more detail in Annex 6, but we provide an overview of their differences below.

Adjusting the predictions based on the regression residuals can only happen at the specific model level. For example, in sewerage this would mean adjusting each of the treatment models, the network model, and each wholesale model by a different percentage based on the upper quartile in each model. However, doing this separately for network and treatment may lead to cherry picking as there may be trade-offs between network and treatment costs. In other words, a company which is very low cost in terms of treatment may have less scope to be low cost in relation to its network. This becomes a more significant issue in relation to combining the advanced regression results with the unit cost models. We therefore do not recommend applying

---

[43] These are the three additional values that Ofwat's RBR benchmarks are based on, though other adjustments are possible using the same method.

the residual-based method at the disaggregated level, even though this might be considered a more theoretically correct approach.[44]

The alternative approach is to calculate the lower quartile of the companies' ratios of actual and predicted costs (corresponding to upper quartile efficiency), as in Equation 5.1.

$$UQ\ adjustment = LQ\ (actual\ costs/estimated\ costs) \qquad (5.1)$$

The upper quartile adjustment is then used as a 'scaling factor' to shift the companies' predicted totex.[45] The advantage of this approach is that it avoids cherry-picking as this adjustment can be made after the predictions from all models have been aggregated. We believe that this approach is also more replicable and transparent than the residual based approach. It does, however, assume time invariant inefficiency across all models as the average is taken across all years.[46]

5.2.2. Ratio approach: historical or forecast efficiencies

A caveat of the ratio based approach is that efficiencies can be calculated using either the actual (historical) costs or the companies' own future forecasted expenditure in the numerator of the equation above. The former compares companies' performance to their historical benchmark performance, while the latter provides a relative comparison at a point in the future (i.e. over AMP6). The implication of this is that by using efficiencies based only on future forecasted expenditure over AMP6 there will be a certain number of companies (at least a quarter) whose cost assessment will result in them meeting their upper quartile target. On the other hand, by using only historical data it is theoretically possible to have any number of companies meet (or fail to meet) their upper quartile target.

Figure 5.3: Historical vs Forecast UQ efficiencies
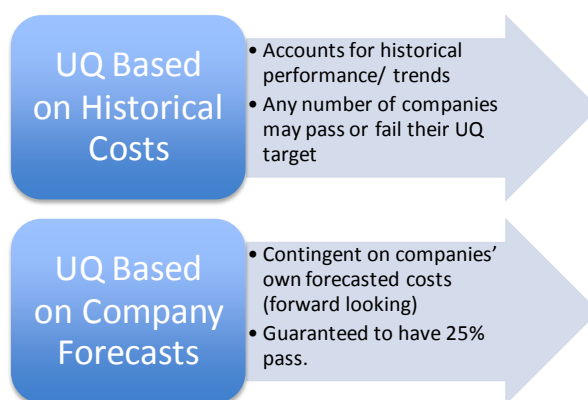
We consider that using the actual expenditure is more consistent with the modelling approach we have adopted and is more independent of the business plan submissions. It is also likely to set a more challenging target as it does not 'guarantee' a certain number of companies will perform better than the upper quartile.

---

[44] In particular, the residual approach would hold the RE models to having time invariant inefficiency.

[45] If we only had one totex model, the two approaches would be the same.

[46] In both the OLS and GLS (RE) models, no decomposition between noise and efficiency is undertaken directly. This adjustment is applied through the use of the upper quartile adjustment.

We therefore recommend using the ratio-based efficiency adjustment with historical costs.

5.2.3. Where to make the adjustment

In the case of the ratio-based approach, the efficiency adjustment can be made at a number of different points in the triangulation diagram without resulting in cherry-picking. The two options we consider most plausible are:

- A. Calculate the upper quartile at the final step of triangulation. That is, triangulate all models and then apply the adjustment.

- B. Calculate the upper quartile at the intermediate stage (i.e. bottom-up and top-down estimates adjusted separately) and then triangulate these intermediate UQ estimates to reach a final UQ estimate.

We illustrate these options in Figure 5.4 below. We applied the same criteria as for selecting the triangulation option. We consider that Option A best meets the criteria as it transparent, and logical and is relatively simple to implement. Moreover, the two options in practice had negligible differences for both water and sewerage.

Figure 5.4: Options for making the UQ adjustment
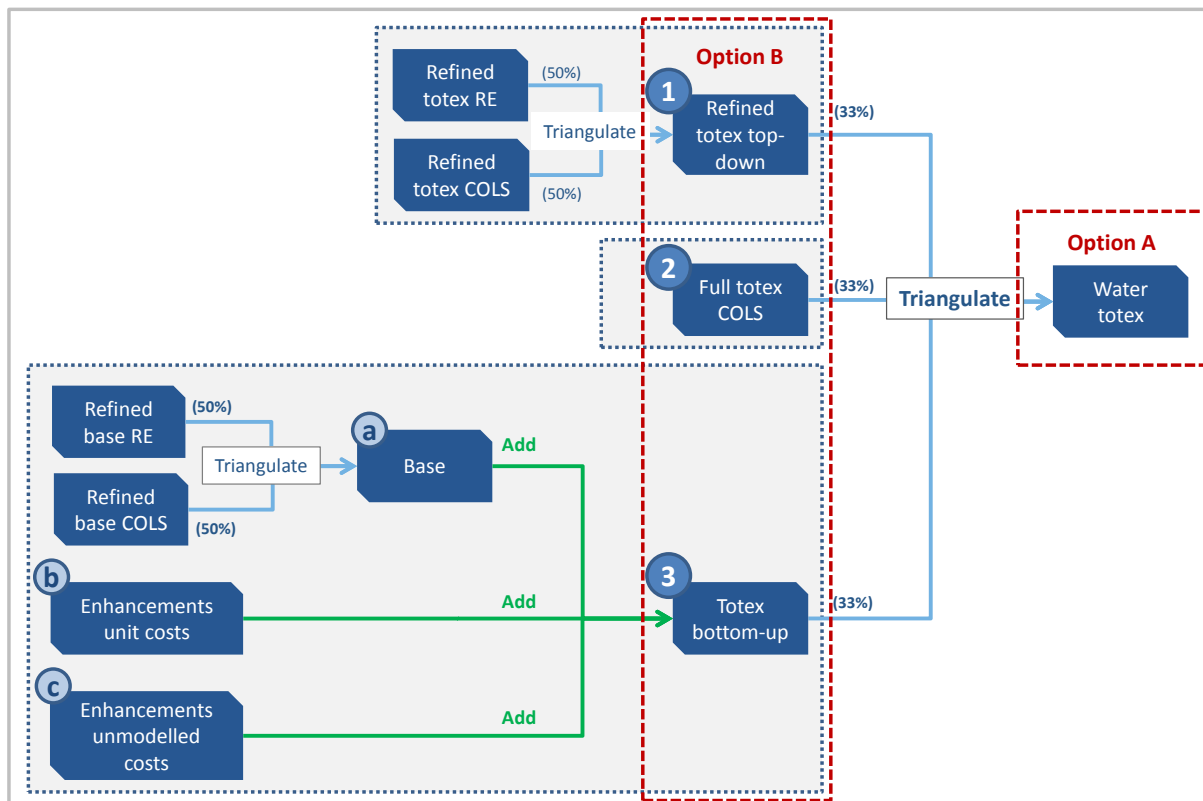
ANNEX 1: EXPLANATORY VARIABLES

A1.1   Water

Most of the variables that we include in our final models are defined in the same way as those we presented in the CEPA *Cost Assessment Report*. However, there are a few additional variables[47] and a few variables that we have defined in a different way[48] in water. Not all of the variables in Table A1.1 are used in every model – the table presents a range and the rationale behind the inclusion of each variable.

Table A1.1: Range of explanatory variables in water models

Type	Variable	Definition	Rationale
Core	Length of mains	Total length of mains at year end	Network scale variable and overall business size proxy
	Property density	Number of connected properties/ length of main	Rural vs. urban divide and economies of density indicator
	Usage*	Potable water/ connected property	Network and resource usage and possible proxy for domestic vs. I&C[49] usage - results similar when normalized by population. *The definition of this variable has changed and it now excludes non-potable water as it is a third party service, for which costs have been excluded.*
	Time trend	Year dummy	Takes into account that the data is for 18 companies over five years and shows the change in costs over the years, including changes in efficiency over time, all other things being equal.
Input prices	Average regional wage*	The data is based on the ONS ASHE SOC surveys by region and allocates companies' service areas to the regions based on Ofwat's updated county allocation. The wages figure is the average hourly salary excluding overtime based on the number of jobs in the company area. The data is transformed to real terms using RPI. Please refer to Annex 3 for more information.	Input price, one of the main cost drivers; the use of these regional indices does not easily deal with the fact that where companies use contractors they may be brought in from other regions and thus have different underlying input prices.

---

[47] Highlighted in blue in the table below.

[48] Marked with an asterisk in the table below.

[49] Industrial and commercial.

Type	Variable	Definition	Rationale
	Regional BCIS index	Provided by Ofwat. The variable uses the construction price index from BCIS, which is based on tender rather than output prices, and allocates the BCIS areas to the companies based on population numbers from the 2001 census. The index was adjusted by the population proportion served within each area. We have used a rolling average in the models where capex is smoothed.	Input price, one of the main capex drivers.[50]
Network characteristics	Population density (occupancy)	Population connected /number of properties connected at year end	Approximates average consumer size (domestic vs. I&C) and can be used to take some of the variation away from usage.
	Proportion of metered properties	(Metered billed households with external meters + metered billed households without external meters + metered billed non-households)/ number of properties connected at year end	Metered customers are assumed to have lower per capita consumption than non-metered customers, thus leading to lower pumping and volume related costs; this variable also captures the wholesale costs related to metering such as installation and replacement. During the period covered, some companies entered the replacement cycle and others had significant increases in meter penetration, which would lead to a positive correlation between proportion of metered properties and totex; it is not clear which factor would be stronger
	Proportion of usage by metered household properties*	Water delivered to billed metered households/(potable water delivered)	In order to estimate the model, one proportion has to be omitted. The omitted variable is non-metered properties and the coefficients on the included variables should be interpreted relative to the one excluded. If the coefficient sign is positive, then metered household properties have higher costs than non-metered properties. *We have updated this variable to reflect the exclusion of non-potable water delivered.*
	Proportion of usage by metered non-household	Water delivered to billed metered non-households/(potable water delivered)	The omitted variable is non-metered properties and the coefficients on the included variables should be interpreted relative to that. Proxy for proportion of

---

Type	Variable	Definition	Rationale
	properties*		I&C customers. *We have updated this variable to reflect the exclusion of non-potable water delivered.*
Treatment and sources characteristics	Sources	Total number of sources/ distribution input	It is a safe assumption that there are economies of scale in the resource and raw water distribution part of the business.
	Pumping head	Pumping head x distribution input	Energy proxy: the higher the pumping head and the lift over which water needs to be pumped, the higher the energy usage – used in old Ofwat opex power model.
	Proportion of water input from river abstractions	Proportion of water input from river abstractions	Proxy for water treatment works (WTW) complexity; boreholes are omitted and considering that boreholes water is generally the cheapest type of source to treat, expect signs to be positive.
	Proportion of water input from reservoirs	Proportion of water input from reservoirs	Same as above
Activity	Proportion of new meters	(selective + optant meters installed)/ (Metered billed households with external meters + metered billed households without external meters + metered billed non-households)	Enhancement activity
	Proportion of new mains	New mains/Total length of mains at year end	Enhancement activity
	Proportion of mains relined and renewed	(mains relined + mains renewed)/ Total length of mains at year end	Maintenance activity
Quality	Properties below reference pressure level	Properties below reference pressure level/total properties connected	Quality measure: the lower the proportion of properties with inadequate water pressure, the higher the costs because companies have spent or are spending money to improve quality but relationship is unclear in the models.
	Leakage	Leakage volume/distribution input	Quality measure: the lower the leakage, the higher the costs because companies have spent money to reduce it; however, companies with a lot of leakage will have to spend more to deal with it – does not always work as quality variable.
	Properties affected by	Properties affected by planned interruptions > 3 hrs/ total	Service quality measure: the more interruptions, the lower the quality; thus

Type	Variable	Definition	Rationale
	unplanned interruptions > 3 hrs	properties connected	if interruptions decrease, this might be associated with service enhancement and thus higher costs, particularly because these interruptions are unplanned.
	Properties affected by planned interruptions > 3 hrs	Properties affected by unplanned interruptions/ total properties connected	Service quality measure: the more interruptions, the lower the quality; thus if interruptions decrease, this might be associated with service enhancement and thus higher costs; planned interruptions however may be correlated with maintenance works and may result in positive sign.

We note that because we have to take the logarithm of the variables and you cannot take the logarithm of zero, we substituted the 0s with 0.001 or 0.00001 depending on whether the variable was a proportion (between 0 and 1) or not.

The correlation coefficients between selected variables listed above are shown in Table A1.2 overleaf. To be clear these are not R-squared values for the correlations. They are the square root of the R-squared values. Highly positive correlations (> 0.5) are highlighted in green, while highly negative correlations (< -0.5) are in orange.

Table A1.2: Correlation between selected water variables

Variable	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
Length of mains (A)	1	-0.02	-0.35	-0.03	-0.18	-0.30	-0.03	-0.10	-0.30	0.89	0.02	-0.15	-0.07	0.48	0.62	0.82	-0.09	-0.05	0.70	0.25	0.00	0.87
Property density (B)	-0.02	1	0.30	0.76	0.63	0.57	0.71	-0.60	-0.42	0.26	-0.47	-0.26	0.23	-0.51	0.00	-0.15	0.16	-0.34	0.22	-0.33	-0.38	0.40
Usage (C)	-0.35	0.30	1	0.46	0.45	0.56	0.25	0.07	-0.12	-0.13	-0.12	0.47	0.32	-0.45	-0.32	-0.39	-0.01	-0.32	-0.34	-0.36	-0.19	-0.08
Average regional wage - entire economy (D)	-0.03	0.76	0.46	1	0.85	0.71	0.73	-0.49	-0.06	0.28	-0.38	-0.32	-0.01	-0.52	-0.06	-0.09	0.13	-0.26	0.15	-0.24	-0.22	0.35
Average regional wage (E)	-0.18	0.63	0.45	0.85	1.00	0.85	0.61	-0.27	0.19	0.08	-0.16	-0.33	-0.11	-0.68	-0.05	-0.16	-0.02	-0.17	-0.12	-0.15	-0.19	0.14
Regional BCIS index (F)	-0.30	0.57	0.56	0.71	0.85	1	0.68	-0.11	0.17	0.01	-0.05	-0.29	-0.02	-0.63	-0.10	-0.34	0.10	-0.28	-0.22	-0.08	-0.13	0.06
Population density (occupancy) (G)	-0.03	0.71	0.25	0.73	0.61	0.68	1	-0.40	-0.03	0.29	-0.18	-0.56	-0.04	-0.47	-0.01	-0.16	0.17	-0.36	0.15	-0.07	-0.17	0.31
Proportion of metered properties (H)	-0.10	-0.60	0.07	-0.49	-0.27	-0.11	-0.40	1	0.31	-0.18	0.92	0.40	0.19	-0.02	0.11	0.03	-0.16	0.21	-0.45	0.22	0.19	-0.28
Number of sources (I)	-0.30	-0.42	-0.12	-0.06	0.19	0.17	-0.03	0.31	1	-0.32	0.39	-0.27	-0.60	-0.27	0.00	-0.07	-0.21	0.25	-0.37	0.12	0.22	-0.42
Pumping head (J)	0.89	0.26	-0.13	0.28	0.08	0.01	0.29	-0.18	-0.32	1	-0.03	-0.22	0.00	0.23	0.52	0.66	0.01	-0.15	0.70	0.22	-0.08	0.94
Proportion of usage by metered household properties (K)	0.02	-0.47	-0.12	-0.38	-0.16	-0.05	-0.18	0.92	0.39	-0.03	1	0.09	0.04	-0.07	0.29	0.16	-0.15	0.20	-0.37	0.28	0.13	-0.15
Proportion of usage by metered non-household properties (L)	-0.15	-0.26	0.47	-0.32	-0.33	-0.29	-0.56	0.40	-0.27	-0.22	0.09	1	0.49	0.00	-0.20	-0.08	-0.15	0.01	-0.23	-0.27	-0.01	-0.19
Proportion of water input from river abstractions (M)	-0.07	0.23	0.32	-0.01	-0.11	-0.02	-0.04	0.19	-0.60	0.00	0.04	0.49	1	-0.12	-0.13	-0.19	0.14	-0.09	-0.04	-0.07	-0.12	0.15

Variable	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
Proportion of water input from reservoirs (N)	0.48	-0.51	-0.45	-0.52	-0.68	-0.63	-0.47	-0.02	-0.27	0.23	-0.07	0.00	-0.12	1	0.20	0.37	0.02	0.04	0.31	0.28	0.11	0.19
Proportion of new meters (O)	0.62	0.00	-0.32	-0.06	-0.05	-0.10	-0.01	0.11	0.00	0.52	0.29	-0.20	-0.13	0.20	1	0.53	-0.09	0.01	0.23	0.23	-0.06	0.51
Proportion of new mains (P)	0.82	-0.15	-0.39	-0.09	-0.16	-0.34	-0.16	0.03	-0.07	0.66	0.16	-0.08	-0.19	0.37	0.53	1	-0.17	0.03	0.50	0.14	0.04	0.62
Proportion of mains renewed or relined (Q)	-0.09	0.16	-0.01	0.13	-0.02	0.10	0.17	-0.16	-0.21	0.01	-0.15	-0.15	0.14	0.02	-0.09	-0.17	1	-0.06	0.14	-0.01	0.47	0.02
Properties below reference pressure level (R)	-0.05	-0.34	-0.32	-0.26	-0.17	-0.28	-0.36	0.21	0.25	-0.15	0.20	0.01	-0.09	0.04	0.01	0.03	-0.06	1	-0.17	0.11	0.34	-0.19
Leakage (S)	0.70	0.22	-0.34	0.15	-0.12	-0.22	0.15	-0.45	-0.37	0.70	-0.37	-0.23	-0.04	0.31	0.23	0.50	0.14	-0.17	1	0.16	0.14	0.74
Properties affected by unplanned interruptions > 3 hrs (T)	0.25	-0.33	-0.36	-0.24	-0.15	-0.08	-0.07	0.22	0.12	0.22	0.28	-0.27	-0.07	0.28	0.23	0.14	-0.01	0.11	0.16	1	0.19	0.13
Properties affected by planned interruptions > 3 hrs (U)	0.00	-0.38	-0.19	-0.22	-0.19	-0.13	-0.17	0.19	0.22	-0.08	0.13	-0.01	-0.12	0.11	-0.06	0.04	0.47	0.34	0.14	0.19	1	-0.10
Distribution input (V)	0.87	0.40	-0.08	0.35	0.14	0.06	0.31	-0.28	-0.42	0.94	-0.15	-0.19	0.15	0.19	0.51	0.62	0.02	-0.19	0.74	0.13	-0.10	1

A1.2 Sewerage

The set of variables that we include in sewerage no longer include any drivers at the sub-company level. They also include additional drivers for treatment and sludge. We have used the same notation to indicate if a variable has been updated or added since the CEPA Cost Assessment Report.

Table A1.3: Range of explanatory variables in sewerage models

Type	Variable	Definition	Rationale
Core	Length of sewers	Total length of sewers at year end	Network scale variable
	Density*	(water and sewerage properties connected+ sewerage only properties connected)/ length of sewers	Rural versus urban divide and another economies of density indicator
	Usage*	Total load entering system/ properties connected[51]	Network usage and possible proxy for domestic versus industrial and commercial (I&C) usage. Since load measures both strength and volume of the sewage that goes into the system and only the volume affects the network costs, it may not be a perfect proxy.
	Time trend	Year dummy	Takes into account that the data is for 10 companies over nine years and shows the change in costs over the years, all other things being equal.
Input prices	Average regional wage*	The data is based on the ONS ASHE SOC surveys by region and allocates companies' service areas to the regions based on Ofwat's updated county allocation. The wages figure is the average hourly salary excluding overtime based on the number of jobs in the company area. The data is transformed to real terms using RPI. Please refer to Annex 3 for more information.	Input price is one of the main cost drivers; assumption is that there is little outsourced outside the region of the company's operation.
	Regional BCIS index	Provided by Ofwat. The variable uses the construction price index from BCIS, which is based on tender rather than output prices, and allocates the BCIS areas to the companies based on population numbers from the 2001 census. The index was adjusted by the population proportion served	Input price is one of the main capex drivers

---

[51] Properties connected include both household and non-households.

Type	Variable	Definition	Rationale
		within each area. The index is originally reported in real terms.	
Network activity	Proportion of sewers replaced and renewed	(Critical sewers replaced + non-critical sewers replaced+ critical sewers renewed + non-critical sewers renewed)/ Total length of sewers at year end	Maintenance activity
Treatment and sludge	Load	Total load in kg BOD_5[52]/day	Size/scale variable and a main cost driver
	Sludge disposed	Total volume ('000 tonnes) of dry solids (ttds)	Size/scale variable and a main cost driver
	Proportion of load in treatment works size bands 1-3	(Load in band 1+ Load in band 2+ Load in band 3)/total load	This variable should be interpreted in reference to proportion of load in the omitted size band, usually band 6. Since Bands 1-3 tend to be more expensive than higher bands in terms of unit costs due to diseconomies of scale, it is expected that a higher proportion of 1-3 load would lead to higher costs.
	Proportion of activated sludge treatment	Load subject to secondary and tertiary activated sludge treatment/Total load	As this is considered the most expensive type of treatment from the ones reported, coefficient sign is expected to be positive. Interpreted against all other treatment type proportion.
	Number of large works with the tight consents dummy	Count of all dummy variables for works with tight consent on suspended solids (SS), Biological Oxygen Demand (BOD_5) and ammonia;[53] 1 if tight consent exists on both SS and BOD_5 or ammonia	As tight consent requires companies to meet certain discharge quality, this will lead to higher opex. It also partially picks up economies of scale.

The correlation coefficients in sewerage are shown in Table A1.4 below.

---

[52] Biological Oxygen Demand.

[53] Thresholds for the determination of consent: 30 mg/l for suspended solids, 20 mg/l for BOD, 5 mg/l for ammonia; if the level of each of these items is below the threshold, tight consent is equal to 1 as it is more expensive achieve a lower or tighter concentration in the consent..

Table A1.4: Correlation between selected sewerage variables

Variable	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Length of sewers (A)	1	-0.09	-0.11	0.71	0.49	0.35	0.96	0.98	-0.31	-0.61	-0.48	-0.53	0.56	0.51	0.95
Property density (B)	-0.09	1	-0.09	0.24	0.39	0.49	0.05	0.07	-0.02	-0.26	-0.42	-0.60	0.48	0.45	0.01
Usage (C)	-0.11	-0.09	1	0.01	-0.28	-0.33	-0.07	-0.04	0.17	-0.33	-0.48	-0.26	0.37	0.22	-0.23
Average regional wage - full economy (D)	0.71	0.24	0.01	1	0.82	0.72	0.80	0.77	-0.31	-0.59	-0.49	-0.59	0.58	0.59	0.63
Average regional wage (E)	0.49	0.39	-0.28	0.82	1	0.83	0.59	0.54	-0.35	-0.39	-0.20	-0.49	0.38	0.39	0.45
Regional BCIS index (F)	0.35	0.49	-0.33	0.72	0.83	1	0.49	0.43	-0.19	-0.25	-0.16	-0.39	0.29	0.47	0.32
Sludge disposed (G)	0.96	0.05	-0.07	0.80	0.59	0.49	1	0.98	-0.31	-0.59	-0.52	-0.58	0.59	0.61	0.90
Total load (H)	0.98	0.07	-0.04	0.77	0.54	0.43	0.98	1	-0.32	-0.67	-0.60	-0.64	0.67	0.61	0.93
Sewers replaced and renovated (I)	-0.31	-0.02	0.17	-0.31	-0.35	-0.19	-0.31	-0.32	1	0.13	0.06	0.05	-0.08	-0.04	-0.30
Proportion of load in treatment works size bands 1-3 (J)	-0.61	-0.26	-0.33	-0.59	-0.39	-0.25	-0.59	-0.67	0.13	1	0.87	0.83	-0.93	-0.49	-0.56
Proportion of load in treatment works size band 4 (K)	-0.48	-0.42	-0.48	-0.49	-0.20	-0.16	-0.52	-0.60	0.06	0.87	1	0.84	-0.95	-0.64	-0.47
Proportion of load in treatment works size band 5 (L)	-0.53	-0.60	-0.26	-0.59	-0.49	-0.39	-0.58	-0.64	0.05	0.83	0.84	1	-0.95	-0.73	-0.53
Proportion of load in treatment works size band 6 (M)	0.56	0.48	0.37	0.58	0.38	0.29	0.59	0.67	-0.08	-0.93	-0.95	-0.95	1	0.68	0.55
Proportion of activated sludge treatment (N)	0.51	0.45	0.22	0.59	0.39	0.47	0.61	0.61	-0.04	-0.49	-0.64	-0.73	0.68	1	0.48
Number of tight consent large works (O)	0.95	0.01	-0.23	0.63	0.45	0.32	0.90	0.93	-0.30	-0.56	-0.47	-0.53	0.55	0.48	1

ANNEX 2: ALTERNATIVE VARIABLES

Besides the variables included in Tables 2.1 and Table 2.2., we considered and tested a range of alternative or additional variables. For the variables we only considered but could not test, data was either not available or was not sufficiently reliable for the time period modelled. Here we mainly discuss the variables that we tried as alternative measures and briefly touch on the ones we would have liked to test.

A2.1   Water

In water, we considered a few variables in addition to those discussed in Table 2.1.

Table A2.1: Alternative variables explored in water

Alternative variable	Original variable	Use	Reason for rejection
Unsmoothed costs	Smoothed costs	Capex profile	Unsmoothed capex is relatively volatile over the years. Models with smoothed capex are more robust.
Quality deltas (change in quality)	Quality variables defined in Table A1.1	Measure quality	Not significant and not acting like a quality variable.
Quality lags	Quality variables defined in Table A1.1	Measure quality	Not significant and not acting like a quality variable.
Pumping head	Pumping head x distribution input	Energy proxy	Identical results
Distribution input	Water delivered	In usage variable	Models with alternative variable did not perform better than original variable.
Non-normalised quality and activity variables	Normalised	Measure quality and activity	Takes away from the core coefficients
Gross weekly average regional wage including all occupations	Hourly average regional wage, weighting two SOC options	Regional wage proxy	Includes a range of occupations not applicable to the water and sewerage industry. We have tried several alternatives here discussed in Annex 3.
Serviceability dummy	N/A	Quality measure	This is not feasible to collect going forward as objectivity is compromised when companies self-assess serviceability.

In addition, we tested various other independent variables in the modelling, but ruled them out because of low data quality and/ or low variance during the given period (even between companies). These variables included:

- refurbished water treatment works (poor data quality);
- number of water treatment works (poor data quality);
- new/replaced water treatment works (poor data quality);
- capacity of water treatment works for maintenance (poor data quality)

- Security Of Supply Index (no variation in data);

- internal floods (sets adverse cost incentives)

- external floods (no data for the entire period); and

- population equivalent (p.e.) of refurbished sewage treatment works (poor data quality).

A2.2 Sewerage

We also considered alternative measures to those set out in Table 2.2, but decided against incorporating them in the preferred sewerage models for various reasons.

Table A2.2: Alternative variables explored in sewerage

Alternative variable	Final variable	Use	Reason for rejection
Unsmoothed costs	Smoothed costs	Capex profile	Unsmoothed capex is relatively volatile over the years. Models with smoothed capex are more robust.
Pumping station capacity	N/A	Proxy for energy use	Only one year of data available. Should be already captured by density and load as it would be correlated with size. No benefit of including it as a constant.
Load entering system / length of sewerage	Load entering system/ properties connected	Usage measure	Highly collinear with density and thus makes the interpretation of coefficients less transparent.
Load entering system

Load treated/ load entering system | Load treated | Load measure in wholesale model | Treatment comprises a larger proportion of the wholesale base costs and thus the model should be more 'treatment'-oriented. Some companies do not have all of their load entering system being treated under the economically regulated entity, so using load entering system (a network variable) for treatment is not appropriate. |
Smoothed wage	Unsmoothed wage	Wages in the water industry are not as volatile year on year as the proxy used (based on household survey in the regions).	Models with alternative variable did not perform better than original variable
Wage index (100 + Ofwat wage differential %)	Unsmoothed wage (level)	To capture regional differences, not over time	Models with alternative variable did not perform better with original variable
Sludge treated	Sludge disposed	Not all sludge treated is disposed	Insufficient data

Alternative variable	Final variable	Use	Reason for rejection
		in that year	
Proportion of load subject to three tight consents	Number of large works with the tight consents dummy	Sludge quality and treatment requirements	Models with alternative variable did not perform better than original variable.
Flooding incidents (overloaded sewers)	N/A	To capture quality of network services, specifically internal flooding	Quality measure: the higher the number of floods, the lower the quality, the lower the cost; as flood incidents are also a source of opex, variable may not work as a quality measure but a maintenance driver
Flooding incidents (overloaded + equipment failure + blockages)	N/A		
Proportion of load in band 4 & Proportion of load in band 5 & Proportion of load in band 6	Proportion of load in bands 1-3	To capture economies of scale in the size of treatment facility	Models with a combination of these were tested. The coefficients were most reasonable when only controlling for bands 1-3. This only changes the interpretation of the coefficients, since by controlling for 1-3 we interpret the coefficients vis-à-vis bands 4-6.
Serviceability	N/A	Quality measure	This is not feasible to collect going forward as objectivity is compromised when companies self-assess serviceability.

ANNEX 3: REGIONAL WAGES

We have tested several options for the wage variable to take into account regional differences in labour costs. All of them are based on data collected from the ONS ASHE survey, allocated to the territory of operation of each company. In our final models, we use a variable that is based on Table 15.6a of the ASHE series, which provides regional hourly earnings by occupation category, excluding overtime pay. This annex describes in detail how we constructed the variable and what alternatives we have tested.

A3.1 Constructing the regional wages variable

ASHE reports the mean and median earnings in its data series. In our analysis, we have decided to use the mean as it better captures the distribution of earnings within the occupation category. In their RIIO-ED1 modelling for Ofgem Frontier Economics tested both mean and median estimates and concluded that the mean was statistically more robust than the median.

We also considered using weekly instead of hourly pay to proxy differences in regional wages. Weekly pay may be capturing differences in company policies and in efficiency. For example, if employees in one company work 40 hours a week while employees in another company work 35 hours a week, doing the same job, this would mean that the weekly wages would allow for that inefficiency. We therefore consider hourly wages to be a better proxy for regional discrepancies outside company control.

We have excluded overtime pay for similar reasons – it may be a better proxy for differences in company policy (in any industry) rather than a proxy for regional differences.

Ideally we would like to use a proxy for regional wage differences in water and sewerage but narrowing it down to the industry we are modelling would lead to an endogeneity problem. If we use the industry specific wage reported in the ASHE (SIC series), the companies have the ability to directly influence that data in the future, and the driver would no longer be outside of their control. On the other hand, we would prefer to capture wages in occupations that are more comparable to water and sewerage sector, rather than using the overall economy differences. Occupations which substantially drive the overall wage differences, particularly in London, such as banking and law would not be representative of water and sewerage and would thus reduce the proxy power of our variable.

We therefore weight together two types of occupations, which we consider are predominant and best capture the regional differences between water and sewerage companies. The ones we exclude and our reasoning behind it are summarised in Table A3.1 below.

Table A3.1: Excluded occupation categories

Category excluded	Reason
Managers and senior officials	We assume at this managerial level there should be a national market
Associate professional and technical occupations	Relevant ones are already covered in the professional occupations
Agriculture, textile, etc	Not relevant

Category excluded	Reason
Sales and customer service occupations	Call centres should be retail
Personal service occupations	Not relevant
Transport and mobile machine drivers and operatives	Not relevant
Elementary occupations	Not relevant

One of the occupations that we include in our wage variable proxies specialist labour, such as engineers, while the other one proxies skilled construction labour. We have used the 2-digit SOC level for each of those:

- Specialist: 21 - Science, research, engineering and technology professionals; and

- Skilled: 53 - Skilled Construction And Building Trades.

We also considered using the 1-digit SOC occupations but that includes occupations that are not applicable to the water and sewerage industry. For example, professional occupations (1-digit alternative for the specialist proxy) includes health workers, teaching professionals, social workers, legal advisers, etc. Although companies may have a few internal lawyers, we assume they do not represent a large proportion even if you look at the wage bill rather than wage level. Data in 3- and more-digit occupation categories are less robust because they rely on smaller sample sizes and may also create industry bias. In terms of the skilled proxy, we also tested a combination of the occupation above (53) and Process, Plant And Machine Operatives (81). The movements across regions are very close to those of only using 53 and we thus think it would not add value to the analysis.

Ofgem used a similar approach in DPCR5 and consulted companies on the relative weighting of these two categories of occupation (specialist and skilled) in the different parts of the distribution value chain. On average they assigned a 60:40 ratio in favour of skilled labour. We checked the implicit weight of these two types of occupations in the ASHE sample and compared it to Ofgem's assumption. We also tested the sensitivity of assigning widely different weights to the two occupations on the overall regional differences. The impact was minor and we have thus stuck with the 60:40 weights in both water and sewerage.

The ASHE SOC data in Table 15.6 provides data at the national and regional level. Unlike the Table 7 series (which is not broken down by occupation), it does not have local area breakdowns. However, Ofwat weighted the local area allocations that it had done internally up to the regional level and we were able to use those regional weights to construct the company specific wage variable. The weights used vary between water and sewerage because the companies often cover different territories. for each service

We note that the specialist proxy occupation category that we use (21) was reported in the ASHE as Science And Technology Professionals prior to 2010-11. These changes would be consistent across regions and companies and we therefore expect that the variable would still be picking up regional differentials. This structural break, however, may result in changes in the interpretation of the time trend and in the significance of the regional wage variable. The pooling test that we conducted does not indicate that the elasticity of cost with respect to wage is significantly different in the two periods (pre and post 2010-11). Since the last year of actual

ASHE SOC data available is 2011-12, we have extrapolated the variable for 2012-13 to be able to use in our analysis. We have assumed that the regional differences have remained the same and that the level of real wages has not changed.

A3.2 Alternative regional wage variables

We tested a few alternative wage variables. All of them were based on the ONS ASHE data.

A3.2.1 Whole economy, local area level

One of them used the gross weekly pay (including overtime) reported in Table 7, which provides a breakdown by local area. The theoretical disadvantages of that variable were:

- The allocation of local areas to companies' territory makes the implicit and bold assumption that if a company requires work to be done, say in Islington, it would hire someone from Islington to do it, rather than someone from its wider region of operation; and

- It encompasses all occupations (including bankers, lawyers, agricultural workers, etc), which would overestimate the differential between rural and urban areas.

This variable resulted in a much wider range of wage levels across companies (around 30%), mainly driven by outlier observations. These are highly unrealistic assumptions, considering water and sewerage workers cover similar activities. This variable is highly correlated with the selected regional wage proxy.

Moreover, using this wage variable affected the time trend, which is much higher if using this alternative wage variable. This could be because the time trend is offsetting the downward trend in the real wage variable constructed this way. Using this variable (and the corresponding high time trend) would then mean assuming that in the future real wages will go down in the same way, which is a bold assumption.

A3.2.2 BCIS-style relative level

We also tested using a wage variable that is similarly constructed to the BCIS variable, i.e. it does not take the changes in wages overtime. In calculating this relative level of wages in each year, the differences in wages would then be picked up in the time trend and values would not be comparable year on year.

The use of this index-like variable (between 0 and 2) resulted in the BCIS and the time trend picking up the wage effect. These models did not have superior predictive power to the ones using the real wage variable described in A3.1. Because of the implications on forecasts and the benefits of capturing the relationship between wages and costs over time, we therefore preferred using the regional real wage variable based on a few occupations, weighted together.

ANNEX 4: WATER TEMPLATES

WM1: Totex full translog RE							
Basic description							
Value chain element	**Expenditure level**	**Capex smoothing**	**Functional form**	**Estimator**	**Data structure**	**Number of observations**	**Number of independent variables**
Water	Totex	5 years	Translog	GLS	Panel, random effects	90	28
Econometric results							

Variable **For those variables whose signs are ambiguous should that be indicated – otherwise it looks like a lot of variables are not expected sign**	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	-0.29124		
Length of mains	.90529***	✓	✓
Density	-0.1413		
Usage	-0.14018		
Length^2	-0.02157		
Density^2	1.18367**	✓	
Usage^2	0.51774		
Length x Density	.66907***	✓	
Length x Usage	0.05205		
Density x Usage	-0.93588		
Time trend	0.00188		✓
Average regional wage	1.23852***	✓	✓
Population density	-0.68133		
Proportion of metered properties	-0.41302		
Sources	-.25322***	✓	
Pumping head	.14322*	✓	✓
Proportion of water input from river abstractions	0.00164		✓
Proportion of water input from reservoirs	-0.01667		
Proportion of new meters	0.01179		✓
Proportion of new mains	-0.02076		
Proportion of mains restored/renovated	.04406***	✓	✓
Properties below reference pressure level	0.00097		

Leakage volume	-0.15273		✓
Properties affected by unplanned interruptions > 3 hrs	0.01949		
Properties affected by planned interruptions > 3 hrs	0.01111		
Proportion of usage by metered household properties	0.24588		✓
Proportion of usage by metered non-household properties	-.28900**	✓	
Regional BCIS	0.1277		✓

Criteria for choosing the best model(s)		
Theoretical correctness	**Functional form seems correct?**	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green Ⓖ
	Efficiency specification	Time-invariant efficiency based on a panel structure. Not a concern because of exhaustive testing of time-varying models that have not proven robust.
Statistical performance	**Sign, size, significance of variables**: Many variables insignificant as well as unexpected sign/magnitude. This is not unexpected given the number of explanatory variables and sample size and is also likely due to multicollinearity. For these reasons we give this model Amber Ⓐ in statistical performance. **Goodness of fit**: 0.996435 (Adjusted R-squared: 0.994798607) **Hausman test:** supports the selection of random effects	
Practical Implementation	**Replicability/ transparency:** Average.	
Robustness testing	**Rank correlations:** Rankings and scores very similar to other full totex models but slightly different from refined models. Scores also differ from the base models which exclude enhancement. **Robustness to specification:** The refinement of this model showed sensitivity of coefficients to dropping variables (likely in part due to multicollinearity). Therefore given Amber Ⓐ. **Pooling test**: Pooling of variables across time was tested. Evidence only of very small and immaterial wage pooling pre/post financial crisis (taken as 2008/09). This finding was seen as less robust because pre-crisis coefficient was based on a single year of data.	

Companies	Cost efficiency	Rank
Anglian Water Services	89.2%	14
Dwr Cymru Cyfyngedig (Welsh)	90.5%	12
Northumbrian Water Ltd	95.3%	3
Severn Trent Water Ltd	100.0%	1
South West Water Ltd	93.8%	6

	Southern Water Services Ltd	93.5%	8
	Thames Water Utilities Ltd	91.7%	11
	United Utilities Water Plc	80.9%	18
	Wessex Water Services Ltd	89.5%	13
	Yorkshire Water Services Ltd	93.7%	7
	Affinity Water	86.6%	16
	Bristol Water plc	84.0%	17
	Dee Valley Water Plc	92.4%	10
	Portsmouth Water Ltd	94.1%	5
	Sembcorp Bournemouth Water	92.9%	9
	South East Water Ltd	94.9%	4
	South Staffordshire Cambridge	96.6%	2
	Sutton & East Surrey Water Ltd	86.9%	15

WM2: Totex translog RE without BCIS							
Basic description							
Value chain element	**Expenditure level**	**Capex smoothing**	**Functional form**	**Estimator**	**Data structure**	**Number of observations**	**Number of independent variables**
Water	Totex	5 years	Translog	GLS	Panel, random effects	90	27
Econometric results							
Variable				**Coefficient**		**Statistically significant**	**Expected sign/ magnitude of coefficient (confidence interval)**
Constant				-0.406			
Length of mains				.89700***		✓	✓
Density				-0.13344			
Usage				-0.14851			
Length^2				-0.02536			
Density^2				1.14365**		✓	
Usage^2				0.56372			
Length x Density				.68183***		✓	
Length x Usage				0.07234			
Density x Usage				-0.9283			
Time trend				0.00176			✓
Average regional wage				1.23810***		✓	✓
Population density				-0.68345			
Proportion of metered properties				-0.36818			
Sources				-.24802***		✓	
Pumping head				.15226**		✓	✓
Proportion of water input from river abstractions				0.00167			✓
Proportion of water input from reservoirs				-.01871*		✓	
Proportion of new meters				0.01122			✓
Proportion of new mains				-0.02081			
Proportion of mains restored/renovated				.04454***		✓	✓
Properties below reference pressure level				0.00102			
Leakage volume				-0.1549			✓

Properties affected by unplanned interruptions > 3 hrs	0.01959			
Properties affected by planned interruptions > 3 hrs	0.01074			
Proportion of usage by metered household properties	0.21248			✓
Proportion of usage by metered non-household properties	-.29520**		✓	

Criteria for choosing the best model(s)		
Theoretical correctness	**Functional form seems correct?**	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green **G**
	Efficiency specification	Time-invariant efficiency based on a panel structure. Not a concern because of exhaustive testing of time-varying models that have not proven robust.
Statistical performance	**Sign, size, significance of variables**: Many variables insignificant as well as unexpected sign/magnitude. This is not unexpected given the number of explanatory variables and sample size and is also likely due to multicollinearity. Therefore given Amber **A** . **Goodness of fit**: 0.996309 (Adjusted R-squared: 0.994701629) **Hausman test:** supports the selection of random effects	
Practical Implementation	**Replicability/ transparency:** Average.	
Robustness testing	**Rank correlations:** Rankings and scores very similar to other full totex models, including Model 1, but slightly different from refined models. Scores also differ from the base models which exclude enhancement. **Robustness to specification:** Refinement of this model showed sensitivity of coefficients to specification. Therefore given Amber **A** . **Pooling test**: Pooling of variables across time was tested. Similar to Model 1, no substantial difference in coefficients over time.	

Companies	Cost efficiency	Rank
Anglian Water Services	89.0%	14
Dwr Cymru Cyfyngedig (Welsh)	90.5%	12
Northumbrian Water Ltd	94.9%	3
Severn Trent Water Ltd	100.0%	1
South West Water Ltd	93.8%	7
Southern Water Services Ltd	93.6%	8
Thames Water Utilities Ltd	91.6%	11
United Utilities Water Plc	80.6%	18
Wessex Water Services Ltd	89.4%	13

	Yorkshire Water Services Ltd	94.1%	5
	Affinity Water	86.5%	16
	Bristol Water plc	83.5%	17
	Dee Valley Water Plc	92.4%	10
	Portsmouth Water Ltd	94.0%	6
	Sembcorp Bournemouth Water	93.0%	9
	South East Water Ltd	94.9%	4
	South Staffordshire Cambridge	97.0%	2
	Sutton & East Surrey Water Ltd	86.6%	15

WM3: Totex full translog COLS without BCIS							
Basic description							
Value chain element	**Expenditure level**	**Capex smoothing**	**Functional form**	**Estimator**	**Data structure**	**Number of observations**	**Number of independent variables**
Water	Totex	5 years	Translog	OLS	Pooled cross-section	90	27
Econometric results							

Variable	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	-0.96128		
Length of mains	.90456***	✓	✓
Density	-0.27601		
Usage	-0.03222		
Length^2	-0.03077		
Density^2	1.15405***	✓	
Usage^2	-0.24695		
Length x Density	.64729***	✓	
Length x Usage	-0.00603		
Density x Usage	-0.06318		
Time trend	0.01193		✓
Average regional wage	1.49168***	✓	
Population density	-0.56056		
Proportion of metered properties	-0.77579		✓
Sources	-.29272***	✓	
Pumping head	0.12203		✓
Proportion of water input from river abstractions	0.00224		✓
Proportion of water input from reservoirs	-0.01501		
Proportion of new meters	0.02846		✓
Proportion of new mains	-.03075**	✓	
Proportion of mains restored/renovated	.02901**	✓	✓
Properties below reference pressure level	0.00295		
Leakage volume	-0.20009		✓

Properties affected by unplanned interruptions > 3 hrs	0.00779			
Properties affected by planned interruptions > 3 hrs	0.02661			
Proportion of usage by metered household properties	0.5006			✓
Proportion of usage by metered non-household properties	-0.17073			

Criteria for choosing the best model(s)		
Theoretical correctness	**Functional form seems correct?**	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green **G**
	Efficiency specification	Time-varying efficiency based on a pooled structure.
Statistical performance	**Sign, size, significance of variables**: Many variables insignificant as well as unexpected sign/magnitude. Density and usage are unexpected signs but insignificant. Therefore given Amber **A** **Goodness of fit**: 0.996785 (Adjusted R-squared: .99546)	
Practical Implementation	**Replicability/ transparency:** Average.	
Robustness testing	**Rank correlations:** Rankings and scores very similar to other full totex models, including Models 1 and 2, but slightly different from refined models. Scores also differ from the base models which exclude enhancement. **Robustness to specification:** Refinement of this model showed sensitivity of coefficients to specification. Therefore given Amber **A** **Pooling test**: Pooling of variables across time was tested. No substantial difference in coefficients over time.	

Companies	**Rebased cost efficiency**	**Ranks**
Anglian Water Services	90.1%	13
Dwr Cymru Cyfyngedig (Welsh)	89.5%	14
Northumbrian Water Ltd	97.7%	2
Severn Trent Water Ltd	100.0%	1
South West Water Ltd	94.9%	3
Southern Water Services Ltd	92.1%	9
Thames Water Utilities Ltd	91.9%	10
United Utilities Water Plc	83.0%	18
Wessex Water Services Ltd	90.9%	12
Yorkshire Water Services Ltd	91.0%	11

	Affinity Water	87.1%	17
	Bristol Water plc	87.3%	16
	Dee Valley Water Plc	92.3%	7
	Portsmouth Water Ltd	94.7%	4
	Sembcorp Bournemouth Water	92.2%	8
	South East Water Ltd	94.5%	5
	South Staffordshire Cambridge	94.4%	6
	Sutton & East Surrey Water Ltd	88.9%	15

WM4: Totex refined CD RE without BCIS							
Basic description							
Value chain element	**Expenditure level**	**Capex smoothing**	**Functional form**	**Estimator**	**Data structure**	**Number of observations**	**Number of independent variables**
Water	Totex	5 years	Cobb-Douglas	GLS	Panel, random effects	90	9

Econometric results			
Variable	**Coefficient**	**Statistically significant**	**Expected sign/ magnitude of coefficient (confidence interval)**
Constant	-8.75224***		
Length of mains	1.11822***	✓	✓
Density	0.09766		✓
Time trend	-0.00329		✓
Average regional wage	1.03174***	✓	✓
Population density	0.90024		✓
Proportion of mains restored/renovated	.05567***	✓	✓
Proportion of water input from river abstractions	0.00892		✓
Proportion of water input from reservoirs	-0.01441		

Criteria for choosing the best model(s)		
Theoretical correctness	**Functional form seems correct?**	Test carried out to see effect of Cobb-Douglas formulation. In general statistical testing favours the translog. Therefore given Amber Ⓐ
	Efficiency specification	Time-invariant efficiency based on a panel structure. Not a concern because of exhaustive testing of time-varying models that have not proven robust.
Statistical performance	**Sign, size, significance of variables**: Generally as expected. Density is lower than expected. Proportion of water from reservoirs is negative. Therefore given Amber Ⓐ **Goodness of fit**: 0.978893(Adjusted R-squared: 0.976515125) **Hausman test:** supports the selection of random effects	
Practical Implementation	**Replicability/ transparency:** Average.	
Robustness	**Rank correlations:** Rankings and scores very different from translog models. Range of	

testing	efficiency scores also relatively high and looks implausible. Therefore given Red **R** **Robustness to specification:** This is a refined model, so relatively stable.

Companies	Cost efficiency	Rank
Anglian Water Services	91.9%	2
Dwr Cymru Cyfyngedig (Welsh)	69.5%	13
Northumbrian Water Ltd	77.8%	9
Severn Trent Water Ltd	85.7%	5
South West Water Ltd	67.4%	15
Southern Water Services Ltd	79.5%	8
Thames Water Utilities Ltd	54.1%	18
United Utilities Water Plc	69.9%	12
Wessex Water Services Ltd	68.2%	14
Yorkshire Water Services Ltd	83.0%	7
Affinity Water	83.1%	6
Bristol Water plc	60.5%	17
Dee Valley Water Plc	66.7%	16
Portsmouth Water Ltd	100.0%	1
Sembcorp Bournemouth Water	77.4%	10
South East Water Ltd	86.8%	3
South Staffordshire Cambridge	86.1%	4
Sutton & East Surrey Water Ltd	73.2%	11

WM5: Totex refined translog COLS without BCIS							
Basic description							
Value chain element	Expenditure level	Capex smoothing	Functional form	Estimator	Data structure	Number of observations	Number of independent variables
Water	Totex	5 years	Translog	OLS	Pooled cross-section	90	12

Econometric results			
Variable	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	2.88752*	✓	
Length of mains	1.07182***	✓	✓
Density	0.21036		✓
Length^2	-0.02259		
Density^2	1.06674**	✓	
Length x Density	.51222***	✓	
Time trend	-0.00675		✓
Average regional wage	0.71957		✓
Population density	0.98924		✓
Proportion of mains relined and renovated	.06502***	✓	✓
Proportion of water input from reservoirs	-0.01397		
Proportion of water input from river abstractions	.02014***	✓	✓

Criteria for choosing the best model(s)		
Theoretical correctness	**Functional form seems correct?**	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green **G**
	Efficiency specification	Time-varying efficiency based on a pooled structure.
Statistical performance	**Sign, size, significance of variables**: Generally as expected. Density is slightly lower than expected. Proportion of water from reservoirs is negative but could be due to multicollinearity. Therefore given Green **G** **Goodness of fit**: 0.990676 (Adjusted R-squared: .98936)	

Practical Implementation	Replicability/ transparency: Average.
Robustness testing	**Rank correlations:** Scores and rankings in line with other refined models but different from full models. **Robustness to specification:** This model is relatively refined and stable. Therefore given Green G **Pooling test**: Pooling of variables across time was tested. No substantial difference in coefficients over time.

Companies	Rebased cost efficiency	Ranks
Anglian Water Services	93.5%	7
Dwr Cymru Cyfyngedig (Welsh)	82.9%	15
Northumbrian Water Ltd	89.9%	11
Severn Trent Water Ltd	94.0%	6
South West Water Ltd	90.2%	10
Southern Water Services Ltd	89.1%	12
Thames Water Utilities Ltd	86.8%	14
United Utilities Water Plc	78.0%	16
Wessex Water Services Ltd	87.9%	13
Yorkshire Water Services Ltd	93.1%	8
Affinity Water	98.3%	3
Bristol Water plc	71.1%	18
Dee Valley Water Plc	95.1%	5
Portsmouth Water Ltd	98.1%	4
Sembcorp Bournemouth Water	90.8%	9
South East Water Ltd	99.2%	2
South Staffordshire Cambridge	100.0%	1
Sutton & East Surrey Water Ltd	75.6%	17

WM6: Totex refined translog RE without BCIS							
Basic description							
Value chain element	Expenditure level	Capex smoothing	Functional form	Estimator	Data structure	Number of observations	Number of independent variables
Water	Totex	5 years	Translog	GLS	Panel, random effects	90	12

Econometric results			
Variable	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	2.51229**	✓	
Length of mains	1.07838***	✓	✓
Density	0.28066		✓
Length^2	-0.01917		
Density^2	.94174*	✓	
Length x Density	.55717***	✓	
Time trend	-0.00319		✓
Average regional wage	.95771***	✓	✓
Population density	0.49497		✓
Proportion of mains restored/renovated	.05565***	✓	✓
Proportion of water input from reservoirs	-0.01229		
Proportion of water input from river abstractions	0.01182		✓

Criteria for choosing the best model(s)		
Theoretical correctness	Functional form seems correct?	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green **G**
	Efficiency specification	Time-invariant efficiency based on a panel structure. Not a concern because of exhaustive testing of time-varying models that have not proven robust.
Statistical performance	**Sign, size, significance of variables**: Generally as expected. Density is slightly lower than expected but still with expected sign. Proportion of water from reservoirs is negative but insignificant. Therefore given Green **G**	

	Goodness of fit: 0.990126 (Adjusted R-squared: 0.988587195)
	Hausman test: supports the selection of random effects
Practical Implementation	Replicability/ transparency: Average.
Robustness testing	Rank correlations: Scores and rankings in line with other refined totex models, though different from the full models. Robustness to specification: Relatively refined and stable. Therefore given Green 🟢 G Pooling test: Pooling of variables across time was tested. No substantial difference in coefficients over time.

Companies	Cost efficiency	Rank
Anglian Water Services	95.1%	5
Dwr Cymru Cyfyngedig (Welsh)	79.2%	16
Northumbrian Water Ltd	89.7%	12
Severn Trent Water Ltd	93.3%	7
South West Water Ltd	87.2%	13
Southern Water Services Ltd	92.2%	9
Thames Water Utilities Ltd	85.6%	14
United Utilities Water Plc	79.6%	15
Wessex Water Services Ltd	92.0%	10
Yorkshire Water Services Ltd	92.8%	8
Affinity Water	95.5%	4
Bristol Water plc	69.5%	18
Dee Valley Water Plc	94.8%	6
Portsmouth Water Ltd	98.8%	2
Sembcorp Bournemouth Water	90.6%	11
South East Water Ltd	100.0%	1
South Staffordshire Cambridge	97.7%	3
Sutton & East Surrey Water Ltd	75.0%	17

WM7: Totex refined translog RE with BCIS							
Basic description							
Value chain element	**Expenditure level**	**Capex smoothing**	**Functional form**	**Estimator**	**Data structure**	**Number of observations**	**Number of independent variables**
Water	Totex	5 years	Translog	GLS	Panel, random effects	90	13

Econometric results			
Variable	**Coefficient**	**Statistically significant**	**Expected sign/ magnitude of coefficient (confidence interval)**
Constant	2.33549**	✓	
Length of mains	1.06992***	✓	✓
Density	0.27595		✓
Length^2	-0.02905		
Density^2	.93629**	✓	
Length x Density	.57342***	✓	
Time trend	-0.00198		✓
Average regional wage	1.00698***	✓	✓
Population density	0.52478		✓
Proportion of mains restored/renovated	.05677***	✓	✓
Proportion of water input from reservoirs	-0.01535		
Proportion of water input from river abstractions	0.01197		✓
Regional BCIS	-0.27099		

Criteria for choosing the best model(s)		
Theoretical correctness	**Functional form seems correct?**	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green **G**
	Efficiency specification	Time-invariant efficiency based on a panel structure. Not a concern because of exhaustive testing of time-varying models that have not proven robust.
Statistical performance	**Sign, size, significance of variables**: Density is lower than expected and proportion of water from reservoirs is negative. In this refined model, BCIS is has a highly unexpected	

	sign and since we consider it to be a core variable, we have given Red R
	Goodness of fit: 0.989852 (Adjusted R-squared: 0.988116158)
	Hausman test: supports the selection of random effects
Practical Implementation	**Replicability/ transparency:** Average.
Robustness testing	**Rank correlations:** Scores and rankings in line with other refined models but different from full models.

Robustness to specification: Robust with regards to dropping BCIS but otherwise relatively refined. Therefore given Green G

Pooling test: Pooling of variables across time was tested. No substantial difference in coefficients over time.

Companies	Cost efficiency	Rank
Anglian Water Services	94.8%	5
Dwr Cymru Cyfyngedig (Welsh)	79.3%	16
Northumbrian Water Ltd	89.8%	12
Severn Trent Water Ltd	93.9%	7
South West Water Ltd	87.9%	13
Southern Water Services Ltd	92.1%	10
Thames Water Utilities Ltd	85.7%	14
United Utilities Water Plc	80.1%	15
Wessex Water Services Ltd	92.5%	9
Yorkshire Water Services Ltd	93.3%	8
Affinity Water	95.5%	4
Bristol Water plc	69.3%	18
Dee Valley Water Plc	94.6%	6
Portsmouth Water Ltd	99.5%	2
Sembcorp Bournemouth Water	91.4%	11
South East Water Ltd	100.0%	1
South Staffordshire Cambridge	99.4%	3
Sutton & East Surrey Water Ltd	74.8%	17

WM8: Base refined translog RE with BCIS							
Basic description							
Value chain element	**Expenditure level**	**Capex smoothing**	**Functional form**	**Estimator**	**Data structure**	**Number of observations**	**Number of independent variables**
Water	Opex+base Capex	5 years	Translog	GLS	Panel, random effects	90	13
Econometric results							

Variable	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	1.62291*		
Length of mains	1.02650***	✓	✓
Density	.39851**	✓	✓
Length^2	0.01183		
Density^2	0.33649		
Length x Density	.45041***	✓	
Time trend	.00993*	✓	✓
Average regional wage	.91856***	✓	✓
Population density	1.09772**	✓	✓
Proportion of mains restored/renovated	.03850***	✓	✓
Proportion of water input from reservoirs	-0.00035		✓
Proportion of water input from river abstractions	0.00361		✓
Regional BCIS	-0.19048		

Criteria for choosing the best model(s)

Theoretical correctness	Functional form seems correct?	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green G
	Efficiency specification	Time-invariant efficiency based on a panel structure. Not a concern because of exhaustive testing of time-varying models that have not proven robust.
Statistical performance	**Sign, size, significance of variables**: BCIS is negative. We consider BCIS to be a core	

	variable, therefore given Red **R**
	Goodness of fit: 0.987432 (Adjusted R-squared: 0.985282211)
	Hausman test: supports the selection of random effects
Practical Implementation	**Replicability/ transparency:** Average.
Robustness testing	**Rank correlations:** Scores and rankings in line with other base models.

Robustness to specification: Robust with regards to dropping BCIS, otherwise relatively stable. Therefore given Green **G**

Pooling test: Pooling of variables across time was tested. No substantial difference in coefficients over time.

Companies	Cost efficiency	Rank
Anglian Water Services	87.0%	4
Dwr Cymru Cyfyngedig (Welsh)	65.3%	18
Northumbrian Water Ltd	82.1%	8
Severn Trent Water Ltd	81.9%	10
South West Water Ltd	89.3%	3
Southern Water Services Ltd	79.6%	14
Thames Water Utilities Ltd	81.9%	9
United Utilities Water Plc	80.5%	12
Wessex Water Services Ltd	80.6%	11
Yorkshire Water Services Ltd	83.6%	7
Affinity Water	76.9%	15
Bristol Water plc	69.2%	16
Dee Valley Water Plc	86.2%	6
Portsmouth Water Ltd	94.3%	2
Sembcorp Bournemouth Water	80.1%	13
South East Water Ltd	100.0%	1
South Staffordshire Cambridge	86.3%	5
Sutton & East Surrey Water Ltd	67.3%	17

WM9: Base refined translog COLS without BCIS							
Basic description							
Value chain element	Expenditure level	Capex smoothing	Functional form	Estimator	Data structure	Number of observations	Number of independent variables
Water	Opex+base Capex	5 years	Translog	OLS	Pooled cross-section	90	12

Econometric results			
Variable	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	2.9165		
Length of mains	1.03714***	✓	✓
Density	0.27499		✓
Length^2	0.01439		
Density^2	0.23994		
Length x Density	.35875*	✓	
Time trend	-0.00077		✓
Average regional wage	0.28008		✓
Population density	2.03158**	✓	
Proportion of mains restored/renovated	.05994**	✓	✓
Proportion of water input from reservoirs	-0.00654		
Proportion of water input from river abstractions	0.00477		✓

Criteria for choosing the best model(s)		
Theoretical correctness	**Functional form seems correct?**	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green (G)
	Efficiency specification	Time-varying efficiency based on a pooled structure.
Statistical performance	**Sign, size, significance of variables**: Generally as expected. Population density slightly high, possibly due to multicollinearity. Therefore given Amber (A) **Goodness of fit**: 0.989328 (Adjusted R-squared: .98782)	

Practical Implementation	Replicability/ transparency: Average.		
Robustness testing	**Rank correlations:** Scores and rankings in line with other base models. **Robustness to specification:** Relatively refined and stable. Therefore given Green **G** **Pooling test:** Pooling of variables across time was tested. No substantial difference in coefficients over time..		

Companies	Rebased cost efficiency	Rank
Anglian Water Services	86.6%	7
Dwr Cymru Cyfyngedig (Welsh)	71.1%	18
Northumbrian Water Ltd	84.8%	9
Severn Trent Water Ltd	87.2%	6
South West Water Ltd	91.8%	3
Southern Water Services Ltd	75.5%	15
Thames Water Utilities Ltd	82.7%	12
United Utilities Water Plc	79.6%	14
Wessex Water Services Ltd	82.1%	13
Yorkshire Water Services Ltd	87.3%	5
Affinity Water	83.5%	10
Bristol Water plc	73.3%	16
Dee Valley Water Plc	86.2%	8
Portsmouth Water Ltd	96.1%	2
Sembcorp Bournemouth Water	83.0%	11
South East Water Ltd	100.0%	1
South Staffordshire Cambridge	90.7%	4
Sutton & East Surrey Water Ltd	71.6%	17

WM10: Base refined translog RE without BCIS							
Basic description							
Value chain element	Expenditure level	Capex smoothing	Functional form	Estimator	Data structure	Number of observations	Number of independent variables
Water	Opex+base Capex	5 years	Translog	GLS	Panel, random effects	90	12

Econometric results			
Variable	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	1.71338*		
Length of mains	1.03225***	✓	✓
Density	.40509**	✓	✓
Length^2	0.01912		
Density^2	0.35379		
Length x Density	.44863***	✓	
Time trend	.00941*	✓	✓
Average regional wage	.90116***	✓	✓
Population density	1.05336**	✓	✓
Proportion of mains restored/renovated	.03764***	✓	✓
Proportion of water input from reservoirs	0.00214		✓
Proportion of water input from river abstractions	0.00388		✓

Criteria for choosing the best model(s)		
Theoretical correctness	**Functional form seems correct?**	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green **G**
	Efficiency specification	Time-invariant efficiency based on a panel structure. Not a concern because of exhaustive testing of time-varying models that have not proven robust.
Statistical performance	**Sign, size, significance of variables**: Generally as expected. Therefore given Green **G**	
	Goodness of fit: 0.987553 (Adjusted R-squared: 0.985613208)	

	Hausman test: supports the selection of random effects.
Practical Implementation	**Replicability/ transparency:** Average.
Robustness testing	**Rank correlations:** Scores and rankings in line with other refined totex models, though different from the full models. **Robustness to specification:** Relatively refined and stable. Therefore given Green **G** **Pooling test**: Pooling of variables across time was tested. No substantial difference in coefficients over time.

Companies	Cost efficiency	Rank
Anglian Water Services	87.0%	4
Dwr Cymru Cyfyngedig (Welsh)	65.1%	18
Northumbrian Water Ltd	82.0%	9
Severn Trent Water Ltd	81.4%	10
South West Water Ltd	89.0%	3
Southern Water Services Ltd	79.8%	13
Thames Water Utilities Ltd	82.2%	8
United Utilities Water Plc	80.1%	12
Wessex Water Services Ltd	80.2%	11
Yorkshire Water Services Ltd	83.2%	7
Affinity Water	76.8%	15
Bristol Water plc	69.3%	16
Dee Valley Water Plc	86.8%	5
Portsmouth Water Ltd	93.6%	2
Sembcorp Bournemouth Water	79.5%	14
South East Water Ltd	100.0%	1
South Staffordshire Cambridge	85.1%	6
Sutton & East Surrey Water Ltd	67.3%	17

ANNEX 5: SEWERAGE TEMPLATES

SW1: Base sewerage network refined translog RE							
Basic description							
Value chain element	**Expenditure level**	**Capex smoothing**	**Functional form**	**Estimator**	**Data structure**	**Number of observations**	**Number of independent variables**
Sewage network	Opex + base capex	7 years	Translog	GLS	Panel, random effects	70 = 10 companies x 7 years	7
Econometric results							

Variable	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	2.38617*		
Length of sewers	.81503***	✓	✓
Density	0.57753		✓
Length^2	0.07573		
Density^2	-2.41709		
Length x Density	-2.80243***	✓	
Time trend	.01923***	✓	✓
Regional wage	0.65243		✓

Criteria for choosing the best model(s)

Theoretical correctness	Functional form seems correct?	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green (G)
	Efficiency specification	Time-invariant efficiency based on a panel structure. Not a concern because of exhaustive testing of time-varying models that have not proven robust.
Statistical performance	**Sign, size, significance of variables**: Generally as expected. Therefore given green (G) **Goodness of fit**: 0.9188448 (Adjusted R-squared: .918614) **Hausman test:** supports the selection of random effects	
Practical Implementation	**Replicability/ transparency:** Average.	

Robustness testing	**Rank correlations:** Scores and rankings in line with the other network model, though different from the full models.		
	Robustness to specification: Refined and stable. Not substantially sensitive to adding marginal variables. Therefore given Green (G)		
	Pooling test: Pooling of variables across time was tested. No evidence of pooling.		
	Company	**Cost efficiency**	**Rank**
	Anglian Water Services	84.9%	5
	Dŵr Cymru Cyfyngedig (Welsh)	64.8%	10
	Northumbrian Water Ltd	93.5%	2
	Severn Trent Water Ltd	78.9%	7
	South West Water Ltd	82.3%	6
	Southern Water Services Ltd	75.4%	8
	Thames Water Utilities Ltd	87.1%	4
	United Utilities Water Plc	69.2%	9
	Wessex Water Services Ltd	92.2%	3
	Yorkshire Water Services Ltd	100.0%	1

SW2: Base sewerage network refined translog COLS							
Basic description							
Value chain element	**Expenditure level**	**Capex smoothing**	**Functional form**	**Estimator**	**Data structure**	**Number of observations**	**Number of independent variables**
Sewage network	Opex + base capex	5 years	Translog	OLS	Pooled cross-section	70 = 10 companies x 7 years	8
Econometric results							

Variable	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	7.54856		
Length of sewers	.93319***	✓	✓
Density	1.68459**	✓	✓
Length^2	.16529*	✓	
Density^2	3.87864		
Length x Density	-2.80880**	✓	
Time trend	-0.00286		✓
Regional wage	-1.11998		

Criteria for choosing the best model(s)

Theoretical correctness	Functional form seems correct?	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green G
	Efficiency specification	Time-varying efficiency based on a pooled structure.

Statistical performance	**Sign, size, significance of variables**: Regional wages are highly negative. Therefore given Red R **Goodness of fit**: 0.9334214 (Adjusted R-squared: .92590)

Practical Implementation	**Replicability/ transparency:** Average.

Robustness testing	**Rank correlations:** Scores and rankings in line with other network models, though different from the full models. **Robustness to specification:** Robust with regards to addition of marginal variables.

Therefore given Green G

Pooling test: Pooling of variables across time was tested. Evidence only of very small and immaterial wage pooling pre/post financial crisis (taken as 2008/09).

Company	Rebased average cost efficiency	Rank
Anglian Water Services	85.4%	4
Dŵr Cymru Cyfyngedig (Welsh)	71.9%	9
Northumbrian Water Ltd	98.2%	2
Severn Trent Water Ltd	84.1%	5
South West Water Ltd	83.1%	6
Southern Water Services Ltd	82.2%	7
Thames Water Utilities Ltd	86.9%	3
United Utilities Water Plc	71.3%	10
Wessex Water Services Ltd	82.1%	8
Yorkshire Water Services Ltd	100.0%	1

SW3: Base sewage treatment and sludge full translog RE							
Basic description							
Value chain element	Expenditure level	Capex smoothing	Functional form	Estimator	Data structure	Number of observations	Number of indepen-dent variables
Sewage treatment + sludge	Opex + base capex	7 years	Translog	GLS	Panel, random effects	70 = 10 companies x 7 years	8

Econometric results			
Variable	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	2.00731		
Load	.83981***	✓	✓
Load^2	0.01338		
Time trend	.02182***	✓	✓
Regional wage	1.21993***	✓	✓
Proportion of load treated by activated sludge	0.06375		✓
Proportion of load treated in bands 1-3	0.15658		✓
Proportion of load treated in bands 4 and 5	-0.01552		
Regional BCIS	-0.33458		

Criteria for choosing the best model(s)		
Theoretical correctness	Functional form seems correct?	Statistical testing favours translog the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). However, one of the key cost drivers (density) not included. Therefore given Amber
	Efficiency specification	Time-invariant efficiency based on a panel structure. Not a concern because of exhaustive testing of time-varying models that have not proven robust.
Statistical performance	**Sign, size, significance of variables**: Regional BCIS is negative, proportion in bands 4 & 5 is negative, and no statistical significance of treatment specific variables (proportion of activated sludge, load treated in bands 1-3, load treated in bands 4&5). We consider BCIS to	

	be a core variable with a very unexpected coefficient, therefore given Red **R**
	Hausman test: supports the selection of random effects
	Goodness of fit: 0.9157674 (Adjusted R-squared: 0.9070181)
Practical Implementation	**Replicability/ transparency:** Average.
Robustness testing	**Rank correlations:** Scores and rankings with little correlation with final chosen models. Very high detected inefficiency (around 40%), which seems implausible. Therefore given Red **R**
	Robustness to specification: Not very robust to refinement or adding density.
	Pooling test: Pooling tests indicate evidence of regional BCIS for the global financial crisis (taken as 2008/09). This caused large movements in BCIS coefficient. Also contributes to the Red traffic light.

Company	Cost efficiency	Rank
Anglian Water Services	79.0%	4
Dŵr Cymru Cyfyngedig (Welsh)	79.3%	3
Northumbrian Water Ltd	73.7%	6
Severn Trent Water Ltd	72.4%	8
South West Water Ltd	73.1%	7
Southern Water Services Ltd	71.5%	9
Thames Water Utilities Ltd	90.5%	2
United Utilities Water Plc	63.4%	10
Wessex Water Services Ltd	100.0%	1
Yorkshire Water Services Ltd	76.1%	5

SW4: Base sewage treatment and sludge CD RE

Basic description

Value chain element	Expenditure level	Capex smoothing	Functional form	Estimator	Data structure	Number of observations	Number of independent variables
Sewage treatment + sludge	Opex + base capex	7 years	Cobb-Douglas	GLS	Panel, random effects	70 = 10 companies x 7 years	6

Econometric results

Variable	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	2.041		
Load	.79982***	✓	✓
Time trend	.02145***	✓	✓
Regional wage	1.18614***	✓	✓
Proportion of load treated by activated sludge	0.08802		✓
Proportion of load treated in bands 1-3	.16168*	✓	✓
Sludge disposed	0.02572		✓

Criteria for choosing the best model(s)

Theoretical correctness	Functional form seems correct?	Test carried out to see effect of Cobb-Douglas formulation. In general statistical testing favours the translog. Also, preferred scale variable (density) not included. Therefore given Amber **A**
	Efficiency specification	Time-invariant efficiency based on a panel structure. Not a concern because of exhaustive testing of time-varying models that have not proven robust.

Statistical performance

Sign, size, significance of variables: Generally as expected. Therefore given Green **G**

Hausman test: supports the selection of random effects

Goodness of fit: 0.9066471 (Adjusted R-squared: 0.8990643)

Practical Implementation

Replicability/ transparency: Average.

Robustness testing

Rank correlations: Scores and rankings with little correlation with final chosen models.

Very high detected inefficiency, therefore given Red **R**

Robustness to specification: Not robust to including translog terms.

Pooling test: Pooling of variables across time was tested. Only small evidence of wage pooling for the start of PR09.

Company	Cost efficiency	Rank
Anglian Water Services	78.8%	3
Dŵr Cymru Cyfyngedig (Welsh)	78.5%	4
Northumbrian Water Ltd	72.6%	7
Severn Trent Water Ltd	69.6%	9
South West Water Ltd	73.6%	6
Southern Water Services Ltd	72.1%	8
Thames Water Utilities Ltd	90.3%	2
United Utilities Water Plc	60.4%	10
Wessex Water Services Ltd	100.0%	1
Yorkshire Water Services Ltd	74.6%	5

SW5: Base sewage treatment and sludge refined translog RE							
Basic description							
Value chain element	Expenditure level	Capex smoothing	Functional form	Estimator	Data structure	Number of observations	Number of independent variables
Sewage treatment + sludge	Opex + base capex	7 years	Translog	GLS	Panel, random effects	70 = 10 companies x 7 years	7

Econometric results

Variable	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	1.27055		
Load	.82780***	✓	✓
Density	-.58885*	✓	✓
Load ^2	0.0846		
Density^2	-2.87877		
Load x Density	-3.59445***	✓	
Time trend	.02331***	✓	✓
Regional wage	1.28032***	✓	✓

Criteria for choosing the best model(s)

Theoretical correctness	Functional form seems correct?	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green (G)
	Efficiency specification	Time-invariant efficiency based on a panel structure. Not a concern because of exhaustive testing of time-varying models that have not proven robust.
Statistical performance		**Sign, size, significance of variables**: Generally as expected. Coefficient on density suggests that more dense more dense areas can take advantage of treatment economies of scale. Therefore given Green (G) **Goodness of fit**: 0.9676082 (Adjusted R-squared: 0.964362) **Hausman test:** supports the selection of random effects
Practical Implementation		**Replicability/ transparency:** Average.
Robustness		**Rank correlations:** Scores and rankings with in line with other preferred model for

testing	treatment. Also similar to wholesale opex+base RE models. Therefore given Green **G**
	Robustness to specification: Not very sensitive to adding non-core variables (such as sludge disposed, proportion treated in small works, etc.).
	Pooling test: Pooling of variables across time was tested. Similar to Model 2, no substantial difference in coefficients over time.

Company	Cost efficiency	Rank
Anglian Water Services	91.6%	8
Dŵr Cymru Cyfyngedig (Welsh)	88.2%	9
Northumbrian Water Ltd	98.1%	3
Severn Trent Water Ltd	94.4%	5
South West Water Ltd	93.9%	6
Southern Water Services Ltd	93.2%	7
Thames Water Utilities Ltd	99.1%	2
United Utilities Water Plc	84.4%	10
Wessex Water Services Ltd	100.0%	1
Yorkshire Water Services Ltd	95.1%	4

SW6: Base sewage treatment and sludge refined translog COLS							
Basic description							
Value chain element	Expenditure level	Capex smoothing	Functional form	Estimator	Data structure	Number of observations	Number of independent variables
Sewage treatment + sludge	Opex + base capex	7 years	Translog	OLS	Pooled cross-section	70 = 10 companies x 7 years	7

Econometric results			
Variable	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	1.70913		
Load	.88110***	✓	✓
Density	-.60886***	✓	✓
Load ^2	.12666***	✓	
Density^2	-2.47179		
Load x Density	-4.51344***	✓	
Time trend	.02146**	✓	✓
Regional wage	1.12747***	✓	✓

Criteria for choosing the best model(s)		
Theoretical correctness	Functional form seems correct?	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green **G**
	Efficiency specification	Time-varying efficiency based on a pooled structure.
Statistical performance	**Sign, size, significance of variables**: Generally as expected, though time trend is a bit high. All variables (save on translog term) are individually statistically significant. Therefore given Green **G** **Goodness of fit**: 0.9712557 (Adjusted R-squared: .96801)	
Practical Implementation	**Replicability/ transparency:** Good.	
Robustness	**Rank correlations:** Scores and rankings with in line with other preferred model for	

testing	treatment. Therefore given Green G . **Robustness to specification:** Not very sensitive to adding non-core variables (such as sludge disposed, proportion treated in small works, etc.). **Pooling test**: Pooling of variables across time was tested. No substantial difference in coefficients over time.

Company	Rebased average cost efficiency	Rank
Anglian Water Services	95.3%	7
Dŵr Cymru Cyfyngedig (Welsh)	89.7%	9
Northumbrian Water Ltd	99.7%	2
Severn Trent Water Ltd	100.0%	1
South West Water Ltd	97.7%	3
Southern Water Services Ltd	96.2%	5
Thames Water Utilities Ltd	97.2%	4
United Utilities Water Plc	87.1%	10
Wessex Water Services Ltd	93.9%	8
Yorkshire Water Services Ltd	95.9%	6

SW7: Base wholesale sewerage full translog RE							
Basic description							
Value chain element	**Expenditure level**	**Capex smoothing**	**Functional form**	**Estimator**	**Data structure**	**Number of observations**	**Number of independent variables**
Sewerage	Opex + base capex	7 years	Translog	GLS	Panel, random effects	70 = 10 companies x 7 years	16
Econometric results							

Variable	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	4.32722***		
Length of sewers	.83460***	✓	✓
Density	1.57994***	✓	
Usage	0.37907		✓
Length^2	0.00474		
Density^2	-3.53657*	✓	
Usage^2	-7.53608***	✓	
Length x Density	-2.99588***	✓	
Density x Usage	8.37491**	✓	
Length x Usage	0.4672		
Time trend	0.0052		✓
Regional wage	0.4889		✓
Proportion of sewers relined and renewed	-0.00531		
Sludge disposed	0.00802		✓
Proportion of load treated by activated sludge	-0.07747		
Proportion of load treated in bands 1-3	.10844*	✓	✓
Number of works with tight consents	0.05407		✓

Criteria for choosing the best model(s)		
Theoretical correctness	**Functional form seems correct?**	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green **G**

	Efficiency specification	Time-invariant efficiency based on a panel structure. Not a concern because of exhaustive testing of time-varying models that have not proven robust.
Statistical performance	**Sign, size, significance of variables**: Proportion of sewers relined and renewed and activated sludge are negative but could be due to multicollinearity. Density is high. Therefore given Amber **A** **Goodness of fit**: 0.9815776 (Adjusted R-squared: 0.9774757) **Hausman test:** supports the selection of random effects	
Practical Implementation	**Replicability/ transparency:** Average.	
Robustness testing	**Rank correlations:** Scores in line with other preferred models for opex+base capex. Rankings similar to Model 9 (RE wholesale) but differ from Model 8 (COLS version of this one). Therefore given Amber **A** **Robustness to specification:** Relatively robust to dropping non-core variables. **Pooling test**: Not conducted as model was further refined.	

Company	Cost efficiency	Rank
Anglian Water Services	92.1%	5
Dŵr Cymru Cyfyngedig (Welsh)	88.8%	9
Northumbrian Water Ltd	91.8%	7
Severn Trent Water Ltd	94.6%	3
South West Water Ltd	91.5%	8
Southern Water Services Ltd	91.9%	6
Thames Water Utilities Ltd	94.2%	4
United Utilities Water Plc	80.9%	10
Wessex Water Services Ltd	96.3%	2
Yorkshire Water Services Ltd	100.0%	1

SW8: Base wholesale sewerage full translog COLS							
Basic description							
Value chain element	**Expenditure level**	**Capex smoothing**	**Functional form**	**Estimator**	**Data structure**	**Number of observations**	**Number of independent variables**
Sewerage	Opex + base capex	7 years	Cobb-Douglas	OLS	Pooled cross-section	70 = 10 companies x 7 years	16

Econometric results

Variable	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	5.43947***		
Length of sewers	.82366***	✓	✓
Density	1.84637***	✓	
Usage	1.04331***	✓	✓
Length^2	-0.01814		
Density^2	-2.81372		
Usage^2	-9.30732***	✓	
Length x Density	-3.65895***	✓	
Density x Usage	13.0344**	✓	
Length x Usage	1.44909***	✓	
Time trend	-0.00113		✓
Regional wage	0.27161		✓
Proportion of sewers relined and renewed	0.01601		✓
Sludge disposed	-0.06083		
Proportion of load treated by activated sludge	0.13232		✓
Proportion of load treated in bands 1-3	.12416***	✓	✓
Number of works with tight consents	.13118*	✓	✓

Criteria for choosing the best model(s)

Theoretical correctness	**Functional form seems correct?**	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero can be rejected, i.e. translog is statistically preferred). Therefore given Green G

	Efficiency specification	Time-varying efficiency based on a pooled structure.
Statistical performance	**Sign, size, significance of variables**: Generally as expected. Density is high, sludge disposed is negative (could be due to multicollinearity). Therefore given Amber **A** **Goodness of fit:** 0.9875257 (Adjusted R-squared**:** .98376)	
Practical Implementation	**Replicability/ transparency:** Average.	
Robustness testing	**Rank correlations:** Scores and rankings in line with preferred OLS model for wholesale opex+base capex (Model 10). Apart from this, scores and rankings show little correlation with Model 7 (full version of this one). Therefore given Amber **A** **Robustness to specification:** Relatively robust to dropping non-core variables. **Pooling test**: Not conducted as model was further refined.	

Company	Rebased average cost efficiency	Rank
Anglian Water Services	92.1%	6
Dŵr Cymru Cyfyngedig (Welsh)	90.2%	9
Northumbrian Water Ltd	92.8%	3
Severn Trent Water Ltd	92.7%	4
South West Water Ltd	94.0%	2
Southern Water Services Ltd	92.3%	5
Thames Water Utilities Ltd	90.8%	8
United Utilities Water Plc	86.4%	10
Wessex Water Services Ltd	91.5%	7
Yorkshire Water Services Ltd	100.0%	1

SW9: Base wholesale sewerage refined translog RE							
Basic description							
Value chain element	Expenditure level	Capex smoothing	Functional form	Estimator	Data structure	Number of observations	Number of independent variables
Sewerage	Opex + base capex	7 years	Translog	GLS	Panel, random effects	70 = 10 companies x 7 years	8

Econometric results

Variable	Coefficient	Statistically significant	Expected sign/ magnitude of coefficient (confidence interval)
Constant	2.48948***	✓	
Density	0.04286		
Load	.88260***	✓	✓
Density^2	-2.64727		
Load^2	0.00753		
Load x Density	-2.06762***	✓	
Time trend	.02429***	✓	✓
Regional wage	1.19874***	✓	✓
Proportion treated in bands 1-3	.15554**	✓	✓

Criteria for choosing the best model(s)

Theoretical correctness	Functional form seems correct?	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green **G**
	Efficiency specification	Time-invariant efficiency based on a panel structure. Not a concern because we have tested an exhaustive amount of time-varying models and they have not proven robust.
Statistical performance	**Sign, size, significance of variables**: Generally as expected. Time trend is a bit high but more in line with treatment than network. Most variables statistically significant. Therefore given Green **G**	
	Goodness of fit: 0.9643004 (Adjusted R-squared: 0.9578228)	
	Hausman test: supports the selection of random effects	
Practical	**Replicability/ transparency:** Average.	

Implementation	
Robustness testing	**Rank correlations:** Scores similar to Model 7 (other RE wholesale model) but rankings change with respect to preferred OLS model for opex+base capex (Model 10). Therefore given Amber A

Robustness to specification: Relatively refined and stable.

Pooling test: Pooling tests carried out on wages. Evidence of inconsequential (very little effect on coefficients) wage pooling for the start of PR09.

Company	Cost efficiency	Rank
Anglian Water Services	83.6%	9
Dŵr Cymru Cyfyngedig (Welsh)	85.4%	6
Northumbrian Water Ltd	86.6%	4
Severn Trent Water Ltd	84.9%	7
South West Water Ltd	85.6%	5
Southern Water Services Ltd	83.8%	8
Thames Water Utilities Ltd	90.6%	3
United Utilities Water Plc	72.1%	10
Wessex Water Services Ltd	100.0%	1
Yorkshire Water Services Ltd	93.7%	2

SW10: Base wholesale sewerage refined translog COLS							
Basic description							
Value chain element	**Expenditure level**	**Capex smoothing**	**Functional form**	**Estimator**	**Data structure**	**Number of observations**	**Number of indepen-dent variables**
Sewerage	Opex + base capex	7 years	Translog	OLS	Pooled cross-section	70 = 10 companies x 7 years	8

Econometric results			
Variable	**Coefficient**	**Statistically significant**	**Expected sign/ magnitude of coefficient (confidence interval)**
Constant	3.39158***	✓	
Density	0.05006		
Load	.97713***	✓	✓
Density^2	-1.21131		
Load^2	.10208***		
Load x Density	-3.78995***	✓	
Time trend	.02006**	✓	✓
Regional wage	.84660**	✓	✓
Proportion treated in bands 1-3	.12711**	✓	✓

Criteria for choosing the best model(s)		
Theoretical correctness	**Functional form seems correct?**	Statistical testing favours the translog (the null hypothesis that the coefficients on the second order terms in the translog are zero is rejected, i.e. translog is statistically preferred). Therefore given Green G
	Efficiency specification	Time-varying efficiency based on a pooled structure.
Statistical performance	**Sign, size, significance of variables**: Generally as expected. Most variables have statistical significance. Therefore given Green G **Goodness of fit:** 0.9578228 (Adjusted R-squared: .97334)	
Practical Implementation	**Replicability/ transparency:** Average.	
Robustness	**Rank correlations:** Scores similar and rankings similar to other wholesale COLS model (Model 8) but rank correlations with respect to preferred RE model (Model 9) rather low.	

testing	Therefore given Amber **A**
	Robustness to specification: Relatively refined and stable.
	Pooling test: Pooling tests carried out on wages. No substantial difference in coefficients over time.

Company	Rebased average cost efficiency	Rank
Anglian Water Services	93.8%	7
Dŵr Cymru Cyfyngedig (Welsh)	90.5%	9
Northumbrian Water Ltd	94.8%	5
Severn Trent Water Ltd	98.9%	2
South West Water Ltd	96.0%	3
Southern Water Services Ltd	94.8%	6
Thames Water Utilities Ltd	95.8%	4
United Utilities Water Plc	81.4%	10
Wessex Water Services Ltd	92.7%	8
Yorkshire Water Services Ltd	100.0%	1

101

ANNEX 6: EFFICIENCY CALCULATIONS AND CHALLENGES

In this section we discuss efficiency calculations and adjustments in more detail. The notions of efficiency and inefficiency are well known in the academic and regulatory literature. Underpinning these concepts is the idea that there exists an efficiency frontier representing best practice, against which all firms may be judged. Inefficiency and in turn efficiency scores are then computed relative to this frontier, with frontier firms obtaining a score of unity.

Whilst the definition of efficiency relative to a frontier is clear, in economic regulation a number of different methodologies have been adopted, each with different assumptions. In turn economic regulators have applied regulatory judgement to the "raw outputs" of cost efficiency models. Below we therefore set out both the assumptions of the models we have adopted, how inefficiency is calculated in those models to generate the "raw output", and also then how those raw outputs might be used to arrive at an appropriate efficiency challenge for the companies for forecasting purposes.

Although in the end we are interested in the efficiency challenge to forecast companies' expenditure, we first explain the methodology for calculating the company efficiency scores in the historical models under the different estimation methods in our final models: GLS (RE) and OLS. This is mostly done for completeness although we also use those efficiency scores to evaluate the robustness of different models under one of our selection criteria.

A6.1 Calculating efficiency

The rationale for calculating efficiencies stems from the assumption about the model error term, namely that the residuals can be decomposed into random noise and inefficiency. In some cases, most notably standard panel model applications (i.e. fixed and random effects models), by making certain assumptions it is possible to obtain efficiency scores without making assumptions regarding the distributions of the noise and inefficiency components. In other cases, in what are described in the literature as stochastic frontier models, it is necessary to make assumptions about the distributions of the two terms in order to decompose the residual (typically it is assumed that the noise term is normally distributed and the inefficiency term takes a "one-sided" half normal distribution). Since these distributional assumptions may be considered arbitrary, other methods which do not rely on those assumptions may be preferred.

As indicated, GLS (RE) and COLS are implemented in different ways and make different assumptions. As such, we calculate the comparative efficiencies for these methods using different methodologies. These are discussed in turn below.

A6.1.1 Random effect models

The RE regression is given by the following equation:

$$y_{it} = \alpha + \sum_{p=1}^{p} \beta_p x_{pit} + \mu_i + v_{it}$$

where α is the constant, β_p are the parameter coefficients of the variables included, μ_i is the time-invariant company effect, and ν_{it} is the error term, which varies across company and time. The residual is then equal to $\mu_i + \nu_{it}$. In the standard panel data literature, this residual is deemed to capture noise and (time invariant) unobserved heterogeneity between companies (in the random effects model the latter is assumed to be uncorrelated with the variables included in the model). This model has however also been applied to give an efficiency interpretation, such that the company effect terms estimated, after a suitable transformation, are interpreted as efficiency scores. As indicated by μ_i, RE assumes that efficiency for a company stays constant over the period modelled.

The standard panel data literature sets out the method for computing the company effects. First, average each company's residuals over time to get to a single average residual value for each company.[54] This can be thought of as giving us the time invariant company effect (which will be given an efficiency interpretation for our purposes), leaving the time varying part of the residual to represent random noise. Here the assumptions that efficiency remains constant over time and that the expected value of the error term is zero are crucial in obtaining efficiency scores.

In the standard panel literature the analysis would stop at that point, with the average residuals having identified the company effects and these effects would not normally be of much interest. However, to go further and obtain efficiency scores from these average residuals, the literature specifies that the company with the minimum average residual is identified, and this corresponds to the most efficient company during the period, i.e. this is the frontier company. This company's efficiency is thus 100% (score of one) and the rest of the companies are benchmarked against it. The efficiencies of the other companies are calculated by subtracting the frontier company's average residual from their individual average residuals and calculating the exponent of the negative of that value. This indicates their position (rank) with respect to the frontier line, i.e. the most efficient company. The averages efficiencies calculated in this way are simply indicative of company rankings and relative positions. The average and other efficiency adjustments that need to be made to the predicted values are discussed in Section A6.2.

A6.1.2 Pooled OLS models (COLS efficiency)

We compute the COLS efficiency scores in a different way due to the assumptions that the pooled OLS models make about the error term. The regression equation that corresponds to this type of model is the following:

$$y_{it} = \alpha + \sum_{p=1}^{K} \beta_p x_{pit} + \varepsilon_{it}$$

where α is the constant, β_p are the parameter coefficients of the variables included, and ε_i is the error term, which varies across company and time. The standard OLS model as used in cost function (as opposed to cost frontier) modelling assumes that there is no inefficiency in the model, with the error term comprising entirely noise. Such models are sometimes referred to in

---

[54] Note that in our models the dependent variable and explanatory variables are specified as natural logarithms of the underlying variables. This has the implication on the arithmetic discussed in this section. For example, subtracting logged values actually reflects division of absolute values.

the literature as average response functions. In the corrected ordinary least squares (COLS) approach, OLS is used to estimate the parameters of the model, but inefficiency is incorporated into the model by adjusting the estimated constant term by shifting the OLS line down so that it passes through the maximum negative residual. With this interpretation, all deviations from the frontier are assumed to be inefficiency (there is no noise). Since this is a strong assumption we explain below how appropriate efficiency targets may be obtained from this model. In this model, efficiency is permitted to vary across firms and over time. The efficiency is allowed to vary over time in a very flexible way, though the assumption that inefficiency varies independently over time could also be questioned. We would normally assume some structure to changes in firm performance over time, i.e. that there is a noise component in the calculated inefficiency. We explain how the derivation addresses these two strong assumptions below, i.e. efficiency varying across time and the error term being interpreted as inefficiency.

To compute the raw efficiency scores from this model, we identify the minimum residual in each individual year and benchmark the companies within each year against these. This allows for different companies to be at the frontier in each year. Under this interpretation the movement in the OLS line (the time trend as calculated in the model) is a change in the average cost of all firms, not a frontier shift, and the frontier shift then is computed using the time trend plus the difference in the minimum residual from year to year. This has the advantage that in the last year the frontier goes through the firm with the lowest cost in that year. To get the average score over the last five years (comparable with RE), we average the efficiency scores of each company across the period.

However, as noted earlier the assumption that all deviations from the frontier represent inefficiency is a very strong assumption. The averaged efficiency scores can then be rebased so that the company with the highest average score becomes 100% efficient. This adjusts the position of the frontier, which would have otherwise been calculated without taking account of noise (with reference to other approaches in the regulatory literature, this approach is similar in nature to the use of, for example, an upper quartile adjustment).

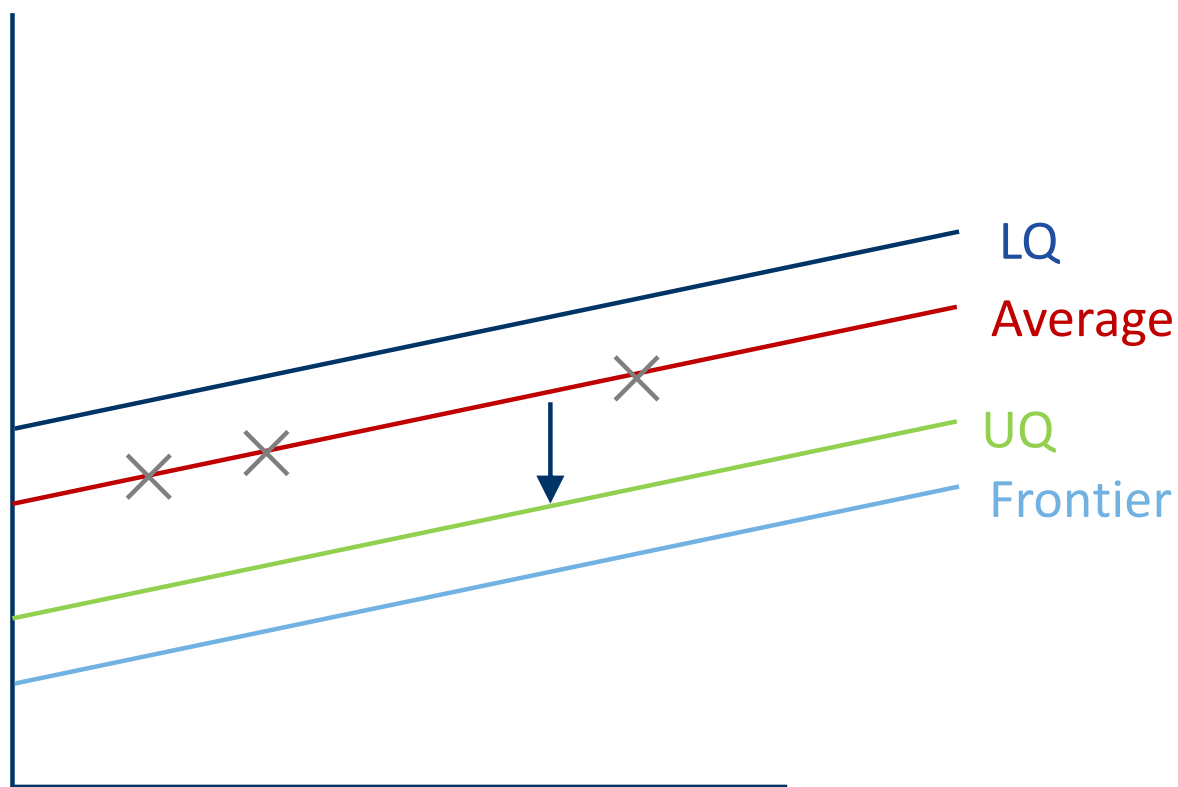## A6.2   Applying efficiency challenges

All the econometric models run for water and sewerage allow for different efficiency challenges to be applied to the predicted expenditures resulting from inserting business plan values for the explanatory variables into the estimated regression equations. The most common options are adjusting companies' costs to the average industry efficiency line, to the frontier company or to the upper quartile industry line. As these lines are identified based on models using historical data, the efficiency challenges applied to the forecasts will reflect the average efficiency level in the previous price control. The estimated regression line includes a time trend and therefore if we apply the coefficient on the time trend (estimated for example based on the last five years of data) we are implicitly assuming that the frontier shift (or ongoing efficiency changes) in the future will be the same in the next five years. It is possible of course for the regulator to impose a different assumption if needed.

The predicted line, $y_{it} = \alpha + \sum_{p=1}^{K} \hat{\beta}_p x_{pit}$, is calculated in different ways under OLS (method used for pooled OLS models) and GLS (used for RE). Therefore each of these models requires a

different method for adjusting the predicted line to yield the same type of efficiency challenged forecast using the approach based on residuals. Another alternative would be to calculate the efficiency adjustment based on the ratio between actual costs and predicted values, which is our preferred option as discussed in Section 5.2. The latter is not based on the model residuals and can be done at a more aggregated level. Residual based approaches would be preferable, but have important drawbacks in terms of feasibility and appropriateness when combining multiple models. As discussed in section 5.2, ratio approaches can only be applied in a consistent manner at the individual model level. This raises issues of potential cherry-picking, replicability, and transparency (discussed in section 5.2).

We discuss the adjustments that would need to be made to the forecasts under the three types of efficiency challenges below. They all essentially result in shifting the average prediction line down by a certain percentage as illustrated in Figure A6.1 below.

Figure A6.1: Illustrating efficiency adjustments

A6.2.1 Average industry line

The term average efficiency refers to applying a challenge to the value predicted by the model that is consistent with a notional company that exhibits the average efficiency of the industry.

That is why, in some cases, such as a **pooled OLS model**, average efficiency forecasts do not require any adjustments to be made to the predicted values as they are based on the average industry line.

This is also the case for **RE**, for which no adjustments need to be made to the predicted values if one wants to predict the average industry efficiency. If, however, Ofwat would like to have a regression line that the average firm in the sample (not the population) lies on, a small adjustment needs to be made to the values predicted. This is needed because although the

average firm in the population is expected to have a residual of zero, this is not the case in the sample and no firm will actually lie on the regression line. We calculate this adjustment for RE models as the difference between the negative of the logged average of the company efficiency scores and the minimum average residual out of all the companies. This can also be expressed in terms of a percentage adjustment to the absolute rather than the logged value. Although we have explained how Ofwat could make this adjustment should they think it appropriate, we recommend that no adjustment is made as the equation will yield the average industry line for a notional average firm.

No adjustment is needed when it comes to the **ratio-based** approach either.

A6.2.2 Frontier company line

Frontier efficiency adjustments reflect the rationale that all companies in an industry are expected to catch up with the most efficient company and should thus be challenged to do so by applying stricter adjustments to the predicted values.

The predicted line in the **RE** model excludes the noise and firm specific effects. Therefore, to get to the frontier, as defined by the best firm in the sample, we need to shift the predicted line down by the frontier company's average residual. We do not recommend using the frontier efficiency for **pooled OLS** models because of the strong no-noise assumption. However, if Ofwat should decide to use frontier efficiency to challenge companies' costs, the predicted values of the pooled OLS model should be adjusted by the average of the minimum residuals in each of the last five years as different companies are allowed to be at the frontier in different years.

The **ratio-based** calculation is done in the following way. We first calculate the efficiency scores of each company by dividing the company's actuals by the estimated value (A). Take the minimum of those efficiencies and apply it to the estimated value (A) to shift the line down to that company.

A6.2.3 Upper quartile industry line

Regulators often use an upper quartile efficiency challenge instead of a frontier challenge as it mitigates the risk of identifying a frontier based on misinterpreting residuals as inefficiency instead of noise. It also sets a more achievable target for companies considering the five-year timeline.

To make an upper quartile adjustment to the predicted values of both **RE** and **pooled OLS** models, we use the upper quartile residual instead of the minimum residual.

As noted in Section A6.1.2, the averaging and rebasing process implied in computing what we referred to as frontier efficiency scores for the pooled OLS models does in fact shift the frontier to some extent. If this method was used it needs to be recognised that application of an upper quartile adjustment in addition to the above approach would result in further deviation from the frontier (though a regulator may consider that such an approach is appropriate based on applying its regulatory judgement).

The **ratio-based** calculation is done in the following way for the UQ. We first calculate the efficiency scores of each company by dividing the company's actuals by the estimated value (A). Take the lower quartile of those ratios (this corresponds to the upper quartile for the industry). Multiply the estimated value (A) by the upper quartile calculated in the previous step.

A6.3 Summary of efficiency adjustments

The table below summarises the approaches taken to adjust the different models for each type of efficiency challenge. They are worked out in a spreadsheet that CEPA has provided to Ofwat separately. In the table, the adjustments described are expressed in terms of changes that need to be done to the logged predicted values but the spreadsheets also provide the % adjustments to the absolute values.

Table A6.1: Efficiency challenge adjustments to the predicted values

	Average	Frontier	Upper quartile
RE	None	Predicted – min (average residuals)	Predicted – upper quartile (average residuals)
Pooled OLS	None	Predicted – average (min residuals over last five years)	Predicted – average (upper quartile residuals over last five years)
Ratio-based	None	$\text{Min}\left(\dfrac{\text{Actual}}{\text{Predicted}}\right)$	lower quartile $\left(\dfrac{\text{Actual}}{\text{Predicted}}\right)$

ANNEX 7: LOGARITHMIC TRANSFORMATION OF PREDICTED VALUES

There are several ways to transform values predicted by log-linear equation into absolute values. Some of the transformation methods require an adjustment to the exponent of the predicted value, while others do not. For the preferred models in this analysis, all transformations have a minor impact on the predicted values because of the sample size. Here we explain the different approaches to log transformation. The rationale behind making an adjustment to the exponent of the log value is that the expected value of the error is zero in logarithmic terms. However, there is no consensus in the academic literature that an adjustment needs to be made, particularly for large samples and financial variables, as argued by William Greene in his textbook *Econometric Analysis*.

The literature on the topic has explored the following approaches to transforming logarithmic data back to costs.

- **Naive estimator**: makes no adjustment to account for the expected value of the logged error being equal to 0.

- **Conditional mean estimator**: makes an adjustment and assumes normal distribution of the errors.

- **Smearing estimator**: does not need to assume normal distribution of the errors. In practice, it yields very similar results to the conditional mean estimator for the sample size that we have.

- **The "alpha factor"** (Ofgem): an adjustment factor that Ofgem used for electricity in 2009 but we have not been able to find supporting literature for it. This is the coefficient of the regression when running the actual cost ($£$m) on the predicted costs ($£$m transformed from logs) without a constant. Ofgem state that this should only be used when the errors are homoscedastic otherwise the correction factor is not constant.[55] For the models in this report, this does not yield results much different from the other two. This factor also assumes normal distribution of the residuals.

The table below summarises the formulae for the different estimators listed above. In these equations, e is the exponent, ε_i is the ith residual, and N is the sample size.

Table A7.1: Log transformation adjustments formulae

Estimator	Adjustment formula
Naive estimator	No adjustment
Conditional mean estimator	$e^{\frac{\sum_{i=1}^{N}\varepsilon_i^2}{2N}}$
Smearing estimator	$\frac{\sum_{i=1}^{N}e^{\varepsilon_i}}{N}$
Alpha factor (Ofgem)	Coefficient of the regression when running the actual cost ($£$m) on the predicted costs ($£$m

---

[55] Ofgem, *RIIO-GD1: Initial proposals – Step-by-step guide for the cost efficiency assessment methodology,* August 2012, page 12.

Estimator	Adjustment formula
	transformed from logs) without a constant

The size of the adjustment decreases as sample size increase. Across all our models, it is very close to 100%. The adjustment factors for the individual water models are presented in Table A7.2 below. It is up to Ofwat to decide which estimator to use.

Table A7.2: Water models adjustment factors

Model	Alpha factor	Smearing estimator	Conditional mean adjustment	Naïve estimator
WM3	99.7%	100.4%	100.2%	100.0%
WM5	101.7%	101.0%	100.5%	100.0%
WM6	101.7%	100.6%	100.6%	100.0%
WM9	100.4%	101.1%	100.4%	100.0%
WM10	99.4%	100.6%	100.6%	100.0%

Table A7.3 below provides the same information for the final sewerage models.

Table A7.3: Sewerage models adjustment factors

Model	Alpha factor	Smearing estimator	Conditional mean adjustment	Naïve estimator
SM1	101.1%	101.0%	101.0%	100.0%
SM5	100.7%	100.4%	100.4%	100.0%
SM6	100.1%	100.4%	100.4%	100.0%
SM9	101.9%	100.5%	100.5%	100.0%
SM10	100.1%	100.3%	100.3%	100.0%

ANNEX 8: NON-NORMALISED COEFFICIENTS OF FINAL MODELS

A8.1   Water

The coefficients presented in Annex 4 are not the ones directly used in Ofwat's feeder models. They are at the sample mean and we have presented them in this way for easy interpretation and to facilitate model comparison during the model selection process. The non-normalised versions of those model results are presented in Table A8.1 below. These are the ones used in Ofwat's feeder models along with Jacobs's exogenous variables to estimate the AMP6 initial threshold. We note that the normalised and the non-normalised models are identical, the only difference is in the presentation of the coefficients of the translog variables (length, density, usage). The rest of the variables should have identical coefficients in both versions. We have presented these results only for the five final models used in triangulation.

Table A8.1: Final water models non-normalised coefficients

Variable	WM3	WM5	WM6	WM9	WM10
Length	3.19829	2.85571	2.91246	1.69157	1.82854
Density	-0.659653	0.74489	-0.280996	-2.0025	-2.16202
Usage	-0.488767				
Length^2	-0.0307684	-0.0225912	-0.0191715	0.0143949	0.0191233
Density^2	1.15405	1.06674	0.94174	0.239944	0.353792
Usage^2	-0.246949				
Length x Density	0.647287	0.512217	0.557174	0.358754	0.448631
Length x Usage	-0.00603146				
Density x Usage	-0.0631846				
Time trend	0.0119295	-0.00674629	-0.0031923	-0.000768856	0.00941448
Regional wage	1.49168	0.719568	0.957711	0.280084	0.901165
Population density	-0.560555	0.989236	0.494968	2.03158	1.05336
Proportion of metered properties	-0.775792				
Sources	-0.292716				
Pumping head	0.122031				
Proportion of water input from river abstractions	0.00224101	0.0201406	0.0118232	0.00477246	0.00387934
Proportion of water input from reservoirs	-0.015007	-0.0139714	-0.0122937	-0.006542	0.00213703
Proportion of new meters	0.0284604				
Proportion of new mains	-0.030748				
Proportion of mains relined and renewed	0.0290132	0.0650153	0.0556453	0.0599445	0.0376357

Variable	WM3	WM5	WM6	WM9	WM10
Properties below reference pressure level	0.0029499				
Leakage volume	-0.200091				
Properties affected by unplanned interruptions > 3 hrs	0.00778847				
Properties affected by planned interruptions > 3 hrs	0.0266115				
Proportion of usage by metered household properties	0.5006				
Proportion of usage by metered non-household properties	-0.170731				
Constant	-22.5662	-15.1977	-17.1336	-12.774	-14.6658

A8.2   Sewerage

Table A8.2 provides the same for the final sewerage models used in triangulation. They are the non-normalised versions of the models shown in Annex 5.

Table A8.2: Final sewerage models non-normalised coefficients

Variable	SM1	SM5	SM6	SM9	SM10
Length	11.2973				
Density	50.4841	70.3423	78.6241	49.3735	59.1464
Load		14.1175	17.0439	9.58371	14.659
Length^2	0.07573				
Denstiy^2	-2.41709	-2.87877	-2.47179	-2.64728	-1.21131
Load^2		0.0846	0.12666	0.00753	0.10208
Length x Density	-2.80243				
Load x Density		-3.59439	-4.51344	-2.0676	-3.78995
Time Trend	0.01923	0.02331	0.02146	0.02429	0.02006
Regional Wage	0.65243	1.28032	1.12747	1.19874	0.8466
Proportion of load treated in Bands 1-3				0.15554	0.12711
Constant	-170.353	-244.736	-281.202	-171.011	-224.343

ANNEX 9: RECOMMENDATIONS FOR PR19

Coming out of the model testing, we consider there are a few areas where Ofwat can collect more data to allow the models to account for elements that we were not able to this time round.

A9.1   Capacity measures

In the dataset currently available to Ofwat, there is no reliable measure of network or treatment capacity (be it in water or in sewerage). In network this would mean taking the diameter as well as the length of the sewers/mains, while in treatment it would reflect the spare capacity of treatment works that could take on additional input/load. The few variables related to capacity that are available from the June Returns seem to be of low data quality. Since a proportion of companies' costs depend on the equipment capacity, we believe that including such variables in the models could further improve the results.

A9.2   Usage measure

In sewerage, the load variable, used in usage takes into account both the volume and strength of the sewage. Since only the volume drives network costs, a better measure of usage would be the one that only takes volume into account (not both).