NERA ECONOMIC CONSULTING

IW IRC2 (2017-18) Assessment – ANNEX Econometric Benchmarking

Prepared for the Commission for Energy Regulation (CER)

May 2016

Project Team

James Grayburn

Edward Day

Adriana Linares

NERA Economic Consulting
Marble Arch House, 66 Seymour Street
London W1H 5BT
United Kingdom
Tel: 44 20 7659 8500 Fax: 44 20 7659 8501
www.nera.com

CONFIDENTIALITY

Contents

List of Tables

List of Figures

1. Introduction

In this annex to the main IRC2 interim review report we describe our approach to econometric benchmarking of IW's proposed operating costs as submitted in its IRC2 submission to the CER in March 2016.

The purpose of this study is to inform our assessment of the efficiency of IW's proposed operating cost expenditure for IRC2.

In developing our modelling approach, we reviewed the models we developed at IRC1 to assess IW's comparative efficiency. We also reviewed the models employed by Ofwat at the most recent price control in England and Wales (PR14), as well as those employed by the Competition and Markets Authority (CMA) in the Bristol Water appeal in 2015. We have also reviewed the models developed by the Utility Regulator for Northern Ireland for Northern Irish Water at PC15. We describe these approaches in detail in this annex.

Our set of models for assessing comparative efficiency at IRC2 draws on these different approaches. Our models also accommodate IW specific circumstances, notably the greater network length per connection on the water side, and the higher unit costs for labour in Ireland compared to UK.

Overall, the opex comparative benchmarking shows that IW's proposed IRC2 expenditure is high compared to a benchmark level of efficient expenditure formed based on UK water and sewerage companies. We consider that a reasonable interpretation of the evidence is that IW's opex costs are at least 70% higher than the long-run efficient level, and potentially around 100% higher, i.e. twice the long-run efficient level.

The structure of this annex is as follows:

- Section 2 summarises our approach to econometric modelling at IRC1 and shows how IW sits in these models on the basis of its updated IRC2 submission;
- Section 3 describes recent evolutions in comparative benchmarking in the UK and in particular Ofwat's approach to benchmarking at PR14 and subsequent CMA critique of Ofwat's models. We discuss potential application of these models to the context of Irish Water and consider limitations;
- Section 4 develops a new set of IRC2 models based on those developed at IRC1, data availability at IRC2 and our review of comparative analysis at PR14;
- Section 0 discusses comparative analysis put forward by Irish Water in response to the CER Q&A process;
- Section 6 draws conclusions, and summarises limitations.

2. Econometric Benchmarking at IRC1

At IRC1, we developed a series of models against which to benchmark IW's operating and capital maintenance cost performance. Our methodology is explained in detail in NERA's technical annex provided to the CER – we provide a summary below.[1] We provide a brief summary of the data and methodology that we employed at IRC1. We also present the results of updating this analysis using the data that IW has submitted at IRC2.

We took all data for the English and Welsh water and sewerage companies (WaSCs) from the annual June Returns submissions to Ofwat, covering the period 2000-01 to 2009-10. For Scottish Water and Northern Ireland Water we used comparable data from annual regulatory accounts.

2.1. Summary of IRC1 Methodology

The purpose of econometric benchmarking is to estimate companies' "inefficiency" – the amount of expenditure that can be avoided, based on comparison with a range of other similar companies. This is done by splitting out company *Actual Expenditure* into three categories, described below and shown in Figure 2.1, and the different estimation techniques differ in how they treat these three categories:

1. *Efficient Expenditure/Frontier Expenditure:* this is the expenditure that a fully efficient company would need to undertake in order to deliver its services. This is found by estimating a "cost function" which relates expenditure to cost drivers (aspects of companies operations that affect its costs).
2. *Inefficient Expenditure:* this is expenditure that is not required for the provision of outputs by a fully efficient company.
3. *Random Noise:* this is expenditure incurred sporadically, often as the result of one-off or infrequent events that are not part of the company's regular ongoing operations.

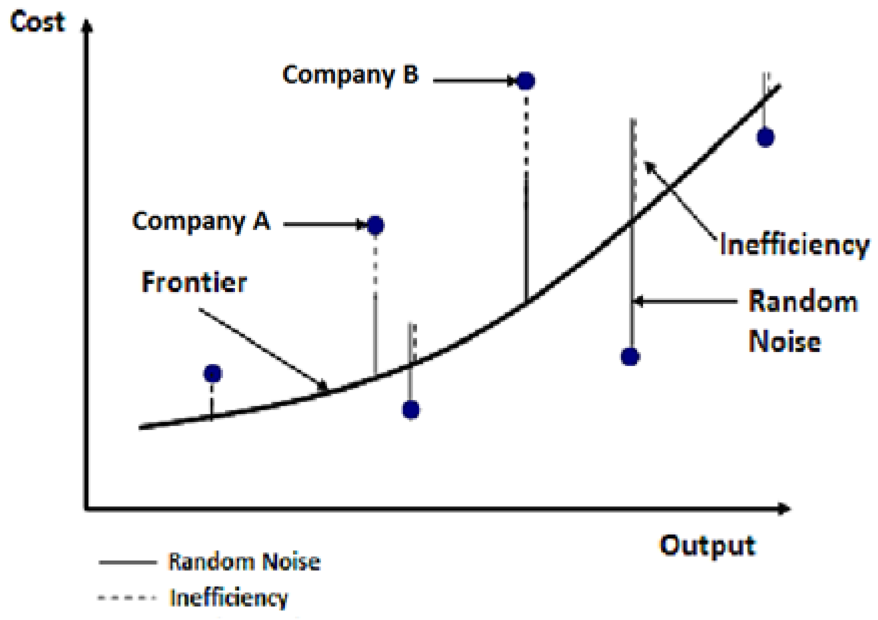
We briefly describe the three estimation techniques used at IRC1 and replicated here (POLS, RE and SFA).[2]

---

[1] NERA (prepared for the Commission for Energy Regulation) (June 2014) *"IW Interim Review Assessment – ANNEX Econometric Benchmarking"*

[2] These stand for Corrected Ordinary Least Squares, Random Effects and Stochastic Frontier Analysis respectively. More description of these techniques, beyond the brief descriptions given in this report, can be found in William Green: "Econometric Analysis", Prentice-Hall International Inc (Section 6 for POLS; Section 14 for RE) and Subal Kumbhakar and Knox Lovell: "Stochastic Frontier Analysis", Cambridge University Press (see Section 3 for SFA model).

Figure 2.1
Efficiency Cost Frontier



Pooled Ordinary Least Squares (POLS) is the simplest and most widely used of the three estimation techniques used at recent reviews (Ofwat PR14; CMA Bristol Water; Ofgem RIIO reviews). POLS estimates *"efficient expenditure"* by estimating the average relationship between expenditure and cost drivers. The POLS technique effectively treats all data observations as though they are from separate companies, even if the data represents the same companies over a period of time. This leads to an estimate of <u>average</u> company expenditure, as well as a "residual" – the difference between this estimate of average expenditure and the company's actual expenditure.

The method then calculates *efficient expenditure* post-modelling. This calculation is done by assuming that:

(a) the most efficient company is the company that has the smallest (i.e. most negative) *residual*;

(b) the most efficient company, identified in (a), has *efficient expenditure* equal to its *actual expenditure.* A percentage adjustment is calculated that is used to adjust the estimated average expenditure from POLS down to its *actual expenditure*;

(c) The adjustment calculated in (b) is used to adjust all the companies' estimated average cost to identify their efficient level of expenditure.

Using POLS, it is difficult to distinguish between *inefficient expenditure* and *random noise* listed above, therefore some proportion is adopted to split the two. For example, the remaining difference between the calculated *efficient expenditure* and *actual expenditure* is assumed to be X% *random noise*, which allows for a split to determine *inefficient expenditure*.

The fact that POLS treats all data observations as individual companies, rather than the reality that they are the same companies over different points in time, is the motivating factor to move to RE or SFA.

Random Effects (RE) models correct the treatment of all data observations as separate companies, by treating data as a "panel" – data which represents a number of companies over a time period. As a result, the estimation of the average relationship between expenditure and cost drivers is likely more accurate. However, like POLS, RE techniques cannot distinguish between inefficient expenditure and random noise which is what motivates the movement to SFA.

Stochastic Frontier Analysis (SFA) is the most sophisticated of the three approaches. Instead of estimating the average relationship between expenditure and cost drivers, SFA explicitly estimates the relationship of *frontier/efficient expenditure* and cost drivers. SFA therefore estimates *inefficient expenditure* during the model estimation, rather than as a post-modelling calculation. *Inefficient expenditure* is produced during model estimation by assuming that it is the element of *actual expenditure* not attributed to *efficient expenditure*, that is consistently positive (it is not possible to have negative inefficiency) and that is present over multiple time periods (inefficiency is not a one-off part of expenditure). However, we note that this approach requires an assumption to be made of the statistical distribution of inefficiency, which may be difficult to objectively define. Additionally, these extra model estimation steps however come at a price, and a large dataset is required to produce accurate results with confidence.

2.2. Water Service Models

Table 2.1 and Table 2.2 show a subset of the models developed to assess IW's proposed water expenditure at IRC1, for each of operating expenditure and capital maintenance respectively. In each case we estimate two models using three estimation techniques (POLS RE and SFA described above), which leads to a total of six models. All models were formed based on the England and Wales companies only.[3]

The tables show the modelled coefficient for each explanatory factor. The tables also show the "p-value" for each factor which indicates whether a variable has a statistically significant effect on costs – a p-value below 0.05 indicates that the factor is statistically significant at the 5% level of significance. The table shows that all explanatory factors in the new models are statistically significant at the 5% level. The tables also show the Adjusted R^2 for each model – a high R^2 indicates that on average a high proportion of expenditure is explained by the regression model.

We were limited in the range of explanatory variables we could include in these models by the availability of data from IW at the time, and make the following observations on the models:

- Our model selection process produces models with similar cost drivers for different expenditure categories, and for different estimation techniques. For example, properties served is the key scale driver for both operating and capital maintenance models.[4]
- Most cost drivers used in the models have intuitive coefficient signs. However, the Water Treatment Works driver has a negative coefficient in some models. We would expect more WTWs to lead to higher costs – a positive effect – although it is possible that a sparsity interpretation of WTWs could imply a negative relationship. The number of WTWs could be viewed as a measure of sparsity, as a water company with a sparse region to serve may have to build more geographically separate water treatment works to serve its population. Sparsity could have a positive or negative impact on costs – more density may imply more congestion and thus more costly maintenance and fault repair; however lower density may imply higher costs as larger distances have to be covered to maintain and fix assets. We include models both with (Model 1) and without (Model 2) the WTWs as an explanatory variable.
- Leakage/DI could be viewed as a measure of asset condition, as a water company with an older inherited network would have poorer asset condition, and therefore suffer more bursts in its network. Although we note this driver could also be explaining other effects such as the explicit costs of fixing leaks.

---

3 Additionally, we are attempting to measure efficient long-term expenditure for IW, which is closer to the steady state levels of expenditure achieved by the E&W companies from 2000 onwards.

4 We only present Opex and CM models in this annex

Table 2.1
Water Opex Models

Model:	Model 1			Model 2		
Estimation Technique:	POLS	RE	SFA	POLS	RE	SFA
Ln(Properties)	1.01 (0.00)	1.03 (0.00)	0.99 (0.00)	0.95 (0.00)	1.01 (0.00)	0.99 (0.00)
Ln(WTW)	-0.22 (0.00	-0.17 (0.00	-0.20 (0.00			
Leakage/DI	1.93 (0.05	1.04 (0.07)	0.78 (0.18)	1.97 (0.07)	0.85 (0.15)	0.67 (0.31)
Intercept	-9.33 (0.00)	-9.33 (0.00)	-9.08 (0.00)	-9.55 (0.00)	-10.09 (0.00)	-9.93 (0.00)
Adjusted R-Squared	0.95	0.95	-	0.94	0.94	-

Source: NERA IRC1 comparative benchmarking

Table 2.2
Water CM Models

Model:	Model 1			Model 2		
Estimation Technique:	POLS	RE	SFA	POLS	RE	SFA
Ln(Properties)	0.72 (0.00)	0.75 (0.00)	0.73 (0.00)	0.75 (0.00)	0.78 (0.00)	0.78 (0.00)
Ln(WTW)	0.10 (0.16)	0.08 (0.35	-0.09 (0.17			
Leakage/DI	1.71 (0.02	1.17 (0.14)	1.32 (0.11)	1.70 (0.04)	1.01 (0.22)	1.03 (0.28)
Intercept	-6.79 (0.00)	-7.00 (0.00)	-1.95 (0.00)	-6.69 (0.00)	-6.99 (0.00)	-7.19 (0.00)
Adjusted R-Squared	0.83	0.83	-	0.82	0.83	-

Source: NERA IRC1 comparative benchmarking

2.3. Sewerage Service Models

Table 2.3 and Table 2.4 show the models developed to assess IW's proposed sewerage service expenditure for opex and CM respectively. As for the water service, all models were formed based on the England and Wales companies only.

We make the following observations on the models:

- Our model selection process produces models with similar cost drivers for all expenditure categories.
- Most cost drivers used in the models have intuitive coefficient signs. However, the Sewer Load driver has a negative coefficient in most models, which is counter-intuitive. As a result, we include models both with and without the Sewer Load driver.

- As for Water Treatment Works, Wastewater Treatment Works could be viewed as a measure of sparsity, as a water company with a sparse region to serve may have to build more geographically separate water treatment works to serve its population. Sparsity could have a positive or negative impact on costs – more density may imply more congestion and thus more costly maintenance and fault repair; however lower density may imply higher costs as larger distances have to be covered to maintain and fix assets.

Table 2.3
Sewerage Opex Models

Model:	**Model 1**			**Model 2**		
Estimation Technique:	POLS	RE	SFA	POLS	RE	SFA
Ln(Properties)	1.23 (0.00)	1.43 (0.00)	1.41 (0.00)	0.79 (0.00)	0.92 (0.00)	0.92 (0.00)
Ln(Sewer Load)	-0.45 (0.28)	-0.58 (0.00)	-0.59 (0.00)			
Ln(WWTW)	0.25 (0.01)	0.28 (0.02)	0.25 (0.00)	0.23 (0.01)	0.25 (0.05)	0.17 (0.00)
Intercept	-8.80 (0.00)	-10.19 (0.00)	-9.86 (0.00)	-7.97 (0.00)	-9.93 (0.00)	-9.68 (0.00)
Adjusted R-Squared	0.93	0.93	-	0.93	0.93	-

Source: NERA IRC1 comparative benchmarking

Table 2.4
Sewerage CM Models

Model:	**Model 1**			**Model 2**		
Estimation Technique:	POLS	RE	SFA	POLS	RE	SFA
Ln(Properties)	0.84 (0.00)	0.88 (0.00)	0.88 (0.00)	1.39 (0.04)	1.98 (0.00)	1.95 (0.00)
Ln(Sewer Load)				-0.55 (0.35)	-1.13 (0.00)	-1.15 (0.01)
Intercept	-7.64 (0.00)	-8.17 (0.00)	-8.40 (0.00)	-8.57 (0.00)	-9.78 (0.00)	-9.36 (0.00)
Adjusted R-Squared	0.75	0.75	-	0.76	0.76	-

Source: NERA IRC1 comparative benchmarking

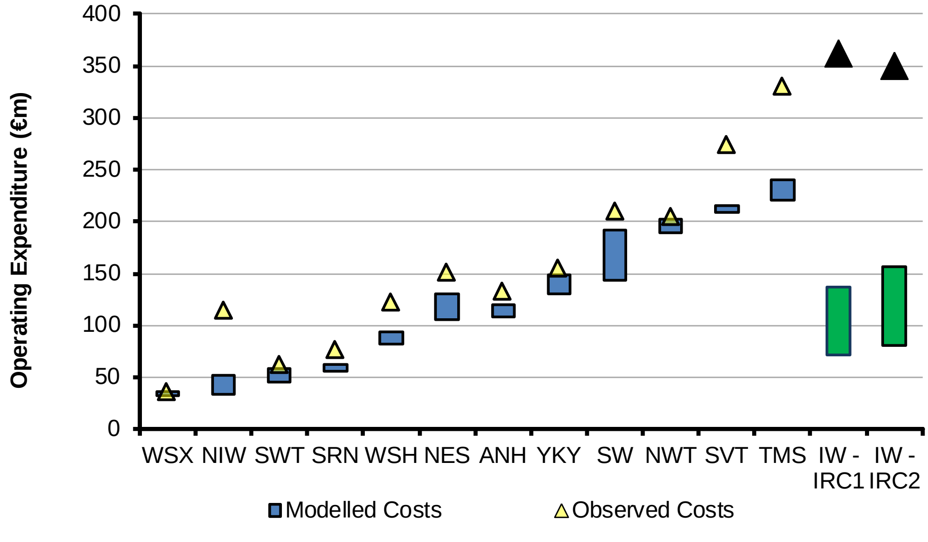
2.4. Results of IRC1 models using updated data for IRC2

We have used the same model result and coefficients obtained from the IRC1 model described above to predict efficient costs for all the UK WaSCs, and for Irish Water. In order to update this analysis to predict modelled costs for IRC2 (2017-18), we have made a number of small adjustments to the models:

- We have updated information on the driver variables used at IRC1 to reflect more recent information provided by Irish Water in its IRC2 submission:
  - The number of connected properties has increased from 1.595 million to 1.706 million for the water service, and from 1.276 million to 1.342 million for the sewerage service.
  - The latest estimate of the number of water treatment works has fallen from 1,027 to 918, while the number of waste water treatment works has risen from 540 to 1,102.
  - Leakage as a proportion of distribution input is now estimated at 43% compared to 41% at IRC1.
  - Load received at sewage treatment works is now estimated to be 328,396 kg BOD/day, while at IRC1 it was assumed to be 187,810 kg BOD/day.
- We have updated all prices to € 2015 prices, using the OECD's Purchasing Power Parities (PPP) for private consumption (we discuss the use of PPP for conversion between currencies in Section 4.3.3). UK company costs have been inflated using RPI CHAW,[5] while Irish Water costs are expressed in € 2015 based on Irish HICP series provided by IW.

Figure 2.2 and Figure 2.3 show the results of this updating exercise. Using the same econometric model as in IRC1, the efficiency gap for the water service has decreased from 165% to 125%, while for the sewerage service it has increased from 164% to 192%.[6]

Figure 2.2
IRC1 Opex Update - Water Modelled Ranges

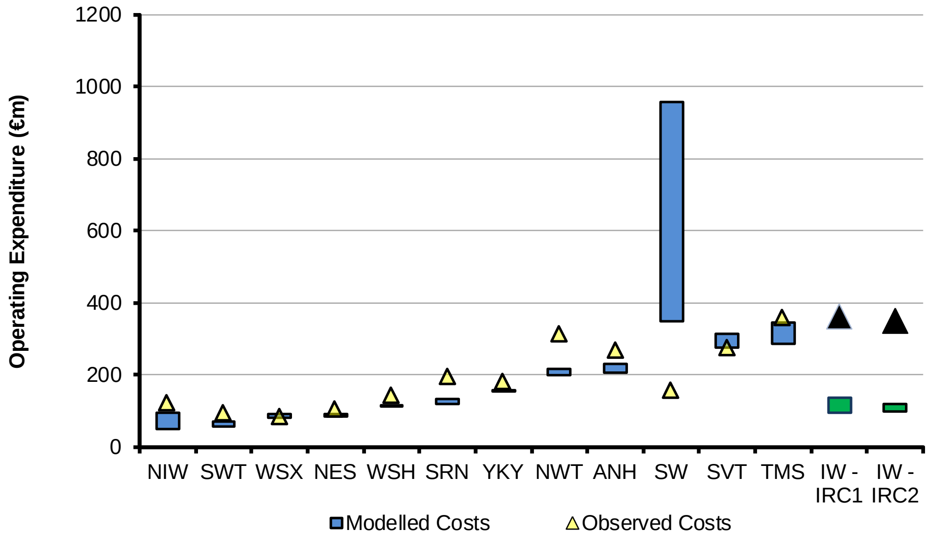


Source: NERA analysis

5 Source: ONS RPI CHAW.

6 Efficiency gap calculated over the upper end of the modelled costs range − i.e. over the model most favourable to IW.

Figure 2.3
IRC1 Opex Update - Sewerage Modelled Ranges



Source: NERA analysis

2.5. Limitations

A key limitation of the IRC1 models was the absence of a large database of explanatory variables from which to "test down" to find the best model specification. We were limited by the availability of driver data available to Irish Water which had only an incomplete understanding of much of the network it had inherited from the 34 local authorities. The limited amount of data precluded more complex model specifications and did not leave much room for sensitivity testing of results.

A second potential drawback of the IRC1 modelling approach is that the main modelling dataset is based on data from England and Wales from 2000-01 to 2009-10, which is by now relatively outdated. We did not consider this a major concern at IRC1, as much of the driver data will not have changed substantially over this period and a ten year panel is long enough to establish a robust relationship between drivers and costs.

Finally, the IRC1 models were developed before subsequent iterations of comparative benchmarking by UK regulators – in particular the Ofgem RIIO-ED1 process and Ofwat's PR14 (and subsequent CMA) analysis.

We address these shortcomings to the extent possible in the following sections, although we note that the data provided by Irish Water at IRC2 is still someway short of what would be expected from a mature utility.

3. Recent Econometric Benchmarking in England and Wales

In this section we present the most recent econometric benchmarking models that have been adopted for the water sector in England and Wales. In Section 3.1 we summarise the model developed by Ofwat for the most recent price review, PR14.

The approach adopted by Ofwat for PR14 has been criticised by the CMA, notably in its assessment of Bristol Water's submission.[7] In light of its reservations on Ofwat's approach, the CMA developed a set of alternative models, which we describe in Section 3.2.

We then explore the possibility of adapting the CMA models to form the basis for comparative analysis of IW's cost proposals at IRC2, presenting a summary of the results of adapted CMA models in Section 3.3 and an analysis of the relevance of these models for the IW IRC2 context in Section 3.4. We conclude that the CMA models are not well suited for adaptation for IRC2, and develop an alternative set of models drawing on UK regulatory precedent in Section 4.

3.1. Ofwat/ CEPA PR14 Model

For the most recent price review – PR14 – Ofwat developed separate sets of econometric models for wholesale water and wholesale sewerage services.[8] The dataset used in these models is publicly available on Ofwat's website.[9] We provide a short summary of the main features of these models below.

For water, Ofwat employs three models to assess Totex (sum of Opex and Capex) and two models to assess Botex (sum of Opex and Capital Maintenance).[10] These five models differ in two dimensions:

- The statistical estimation method**,** which is either Pooled Ordinary Least Squares (POLS) or Random Effects (RE);
- The explanatory factors used. Ofwat uses an "un-refined" model, where it includes all explanatory factors it considers to be relevant; and a "refined" model, where it excludes explanatory factors that are shown not to be statistically important or have an unintuitive effect on costs. The explanatory factors are listed in Table 3.1, and apply to both Totex and Botex models.

The way these two dimensions are combined is illustrated in Figure 3.1, where each set of two boxes represents a model that Ofwat uses.

---

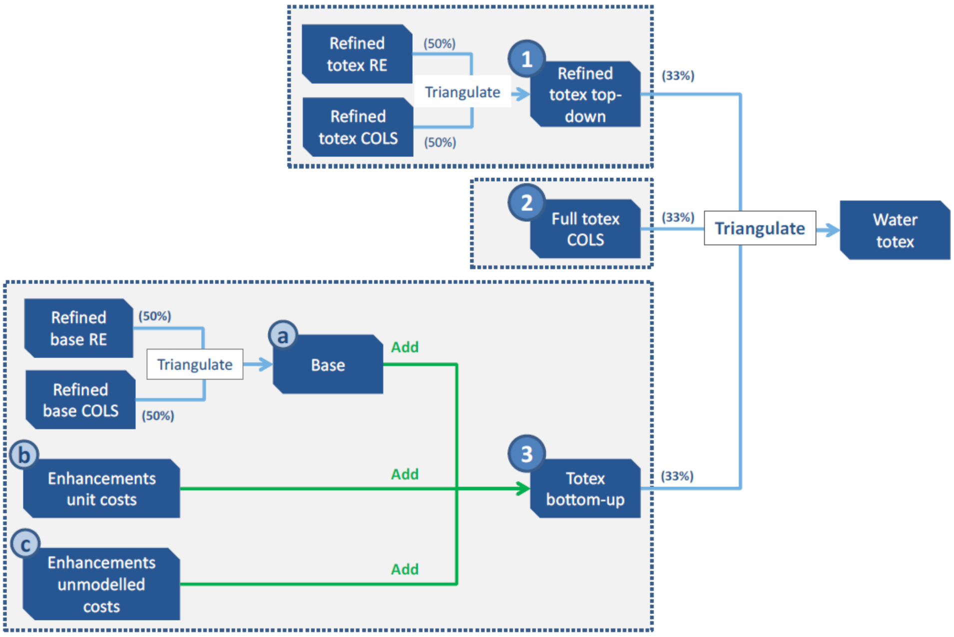
[7] CMA (2015) *"Bristol Water plc - A reference under section 12(3)(a) of the Water Industry Act 1991"*

[8] Ofwat (2014) *"Cost Assessment – Advanced Econometric Models"*. Final report submitted by CEPA.

[9] Ofwat (2014) *"Basic Cost Threshold Feeder Model"*, Appendix A – Water/Wastewater Data Inputs. Available at http://webarchive.nationalarchives.gov.uk/20150624091829/https:/www.ofwat.gov.uk/pricereview/pr14/wholesale/prs_web140404pr14wholesalecostasses

[10] Water botex models are then combined with unit cost models that assess capital enhancement expenditure in order to generate additional measures of totex.

Figure 3.1
Estimation Methods and Sets of Explanatory Variables Used in Ofwat Models – Water Service



Source: NERA Analysis of Ofwat/CEPA (2014) "Cost Assessment – Advanced Econometric Models".

For wastewater, Ofwat similarly combines a range of models, as shown in Figure 3.2. All the models are refined and, as for the water service, different statistical estimation methods (POLS and RE) are combined.

In this case, Ofwat does not consider the Totex approach to be viable, mainly under the criteria of plausibility of the results, and the statistical significance of the estimated coefficients.[11] Instead, it bases all its econometric assessments on Botex.[12]

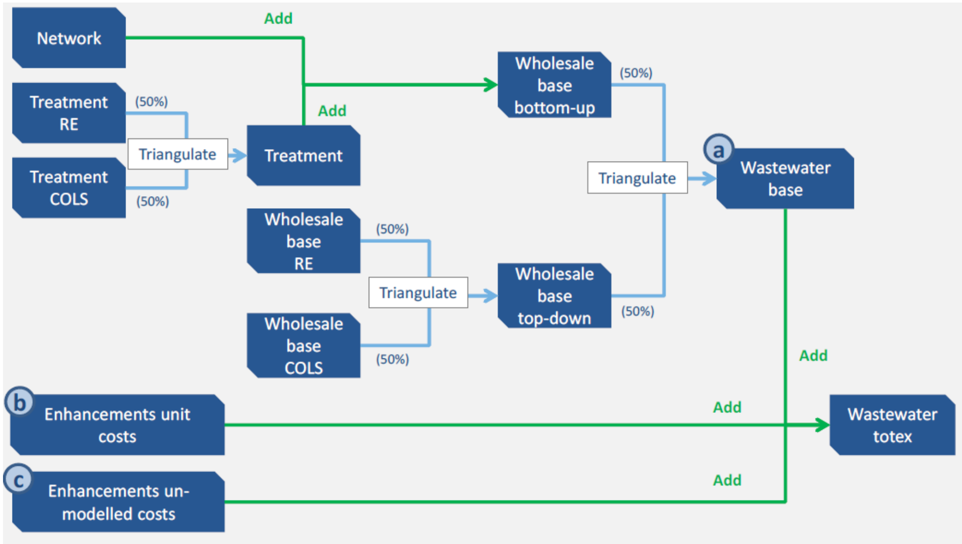
Ofwat uses subservice (Treatment and Network) models as well as aggregate service-level Botex models. The explanatory factors used for each subservice are listed in Table 3.2.

---

[11] CEPA (2013) "*Ofwat Cost Assessment*", pp. 24-25.

[12] Ofwat uses unit cost models to assess capital enhancement expenditure in the sewerage service.

Figure 3.2
Estimation Methods and Sets of Explanatory Variables Used in Ofwat Models – Wastewater Service



Source: NERA Analysis of Ofwat/CEPA (2014) "Cost Assessment – Advanced Econometric Models".

Table 3.1
Explanatory Variables Used by Ofwat in PR14 - Water Service

Description	Inclusion in "un-refined" model	Inclusion in "refined model"
Constant term	✓	✓
Ln (length of mains)	✓	✓
Ln (connected properties / length of mains)	✓	✓
Ln (potable water delivered / connected properties)	✓	
[Ln (length of mains)] ^ 2	✓	✓
[Ln (connected properties / length of mains] ^ 2	✓	✓
[Ln (potable water delivered / connected properties)] ^ 2	✓	
Ln (length of mains) * Ln (connected properties / length of mains)	✓	✓
Ln (length of mains) * Ln (potable water delivered / connected properties)	✓	
Ln (connected properties / length of mains) *Ln (potable water delivered / connected properties)	✓	
Ln (regional wage)	✓	✓
Ln (population supplied / connected properties)	✓	✓
Ln (proportion of properties that are metered)	✓	
Ln (total number of sources / total distribution input)	✓	
Ln (average pumping head * total distribution input)	✓	
Ln (proportion of distribution input from river abstractions)	✓	✓
Ln (proportion of distribution input from reservoirs)	✓	✓
Ln (number of new meters installed in year as a proportion of metered customers)	✓	
Ln (length of new mains laid in year / total length of mains at year end)	✓	
Ln (length of mains relined and renewed / total length of mains at year end)	✓	✓
Ln (number of properties below reference pressure level/total properties connected)	✓	
Ln (volume of leakage / total distribution input)	✓	
Ln (number of properties affected by unplanned interruptions > 3 hrs / total properties connected)	✓	
Ln (number of properties affected by planned interruptions > 3 hrs / total properties connected)	✓	
Ln (potable water delivered to billed metered households / total potable water delivered)	✓	
Ln (potable water delivered to billed metered non-households / total potable water delivered)	✓	
Time trend	✓	✓

Source: NERA Adaptation of CMA Bristol Water Final Determination, Appendix 4.

Table 3.2
Explanatory Variables Used by Ofwat in PR14 - Sewerage Service

Description	Inclusion in "Network Only" model	Inclusion in "Treatment Only" model	Inclusion in "Network + Treatment" model
Constant term	✓	✓	✓
Ln (sewers length)	✓		
Ln (sewage load)		✓	✓
Ln (connected properties / sewers length)	✓	✓	✓
[Ln (sewers length)] ^ 2	✓		
[Ln (sewage load)] ^ 2		✓	✓
[Ln (connected properties / sewers length)] ^ 2	✓	✓	✓
Ln (sewers length) * Ln (connected properties / sewers length)	✓		
Ln (sewage load) * Ln (connected properties / sewers length)		✓	✓
Ln (proportion of load treated in Bands 1-3)			✓
Ln (regional wage)	✓	✓	✓
Time trend	✓	✓	✓

Source: NERA Analysis of Ofwat/CEPA (2014) "Cost Assessment – Advanced Econometric Models".

In its final determination for the Bristol Water case, the CMA identified a number of concerns with the model specifications used by Ofwat: [13]

- **Counter-intuitive coefficients**. Some of the estimated coefficients implied relationships between costs and the explanatory variables that suggested a lack of precision in model estimation and limitations in these models. Ofwat's consultant CEPA identified that the results from these models differed from what it had expected, in terms of both the sign (positive or negative) and magnitude of a number of the estimated coefficients.
- **Number of explanatory variables relative to sample size**. The CMA indicated that the small sample size of Ofwat's models, combined with a large number of explanatory variables – some of which were highly correlated with each other and showed little variation over time – contributed to risks of inaccuracy in the results.

[13] CMA (2015) *"Bristol Water plc - A reference under section 12(3)(a) of the Water Industry Act 1991"*, par. 4.50.

- **Translog models**. Ofwat used models with a particularly complex model specification, which it described as translog.[14] The CMA considered that the models involved relatively complex explanatory variables and it was difficult to interpret the relationships that they implied between costs and explanatory variables in economic or engineering terms. In fact, the inclusion of these variables seemed to have compromised the results – e.g. Ofwat's refined base expenditure models implied a form of diseconomies of scale with respect to the size of a company's customer base, which the CMA found to be counter-intuitive.
- **Relationships between costs and cost drivers**. In some cases, the CMA found Ofwat's models to be specified in a way that implied a relationship between expenditure and a cost driver that did not make sense – e.g. taking logarithms of variables expressed as proportions.
- **Endogeneity**. Some of the explanatory variables used in Ofwat's models of totex and base expenditure represent factors that are, at least in part, under the control of a company's management and cannot be treated as entirely independent of the dependent variable in the model – e.g. explanatory variables for mains renewal, leakage or various quality of service measures.

  The results from a statistical analysis of companies' costs and efficiency may be distorted if some of the explanatory variables were themselves reflective of each company's efficiency and working practices. However, the CMA considered that, given limitations in the available data, it may be better, in some cases, to include an explanatory variable that carries risks of endogeneity than to fail to take any account of potentially important differences between companies.

For the reasons outlined above, the CMA decided to develop a set of alternative models in its final determination for the Bristol Water case. We describe these alternative models, and how we considered adapting them for use in IRC2, below. We note that, for the reasons outlined in Section 3.4, we decided not to include these models in the current price review, and develop our own set of models instead.

3.2. CMA Bristol Alternative Models

The CMA considered an initial set of 18 models for the development of the alternative econometric models for the Bristol Water Final Determination. The models were aimed at estimating the efficient wholesale base expenditure requirements for Bristol Water in PR14.[15]

The specification of the 18 models was based on the combination of the different modelling options described in Table 3.3 and Table 3.4. The final results used in their determination corresponded to the average of the results from a set of seven preferred models, including

[14] See e.g. the variable *Ln (length of mains) * Ln (connected properties / length of mains)* in Table 3.1.

[15] In the CMA models, as in Ofwat PR14 models, base expenditure is defined as the sum of operating expenditure and capital maintenance.

logarithmic and linear unit cost models, and different sets of explanatory variables EV2 and EV3.

As described in Table 3.3, the CMA estimates botex models. Botex represents operating expenditure plus capital maintenance expenditure. We discuss the relevance of these models to the context facing Irish Water at IRC2 below.

Table 3.3
Dimensions in the Specification of the Initial Set of CMA Models

Dimensions	Options Explored in the Initial Set
Model Form	• Logarithmic unit cost models, where the dependent variable is *ln (botex/connected properties)*. • Linear unit cost models, where the dependent variable is *botex/properties*. • Logarithmic aggregate cost models, where the dependent variable is *ln (botex)*.
Dependent Variable	• Smoothed botex (5-year), where botex each year is *(opex in that year)* + *(capital maintenance moving average over five-year period)*. Uses same five-year data sample as Ofwat. • Unsmoothed botex (7-year), where botex each year is *(opex in that year)* + *(capital maintenance in that year)*. Uses longer data period than Ofwat (additional dataset not publicly available).
Explanatory Variables	• Three explanatory variable groups: EV1, EV2 and EV3, described in Table 3.4. • In addition, each model included a constant term and a series of time dummy variables with 2013 as reference year.
Estimation Technique	• Pooled OLS technique.

Source: NERA Summary of CMA Bristol Final Determination.

Table 3.4
Groups of Explanatory Variables Used in CMA Models

Group	Logarithmic unit cost model	Linear unit cost model	Logarithmic aggregate cost models
EV1	• Ln (water delivered per property) • Ln (regional wage measure) • Ln (mains length per property) • Proportion of distribution input from rivers • Proportion of distribution input from reservoirs • Ln (average pumping head)	• Water delivered per property • Regional wage measure • Mains length per property • Proportion of distribution input from rivers multiplied by water delivered per property • Proportion of distribution input from reservoirs multiplied by water delivered per property • Average pumping head multiplied by water delivered per property	• Ln (water delivered per property) • Ln (regional wage measure) • Ln (total mains length) • Ln (total connected properties divided by total mains length) • Proportion of distribution input from rivers • Proportion of distribution input from reservoirs • Ln (average pumping head)
EV2	As for EV1 plus: • Proportion of water consumption by metered non-household customers	As for EV1 plus: • Proportion of water consumption by metered non-household customers	As for EV1 plus: • Proportion of water consumption by metered non-household customers
EV3	As for EV2 but with rivers and reservoirs variables removed and replaced by: • Proportion of distribution input subject to W3 or W4 treatment	As for EV2 but with rivers and reservoirs variables removed and replaced by: • Proportion of distribution input subject to W3 or W4 treatment multiplied by water delivered per property	As for EV2 but with rivers and reservoirs variables removed and replaced by: • Proportion of distribution input subject to W3 or W4 treatment

Source: NERA Summary of CMA Bristol Final Determination.

In order to estimate IW's comparative efficiency, we re-estimate CMA's models as opposed to drawing on CMA's published modelling results and coefficients. We need to re-estimate as the CMA models are based on capital maintenance and operating expenditure, whereas our objective is to develop a model for operating expenditure alone. Also, we do not have entire set of explanatory data included in CMA's models for IW in all cases.

We replicate a selection of the models described in Table 3.3 and Table 3.4. Specifically we estimate:

- One of each of the three model types: logarithmic unit cost model, linear unit cost model and logarithmic aggregate cost model;
- Each of the three model types models using smoothed botex (5-year) as the dependent variable and with the variable group EV2 as explanatory variables;[16]

[16] The CMA's set of preferred models includes only models with explanatory variables groups EV2 and EV3. The group EV3 only differs from EV2 by the use of the variable *"Proportion of distribution input subject to W3 or W4 treatment"*. The data used by the CMA for this variable implies rolling forward the latest available data from Ofwat, which dates from 2009. Given the concerns acknowledged by the CMA about the risks of inaccuracy of these data, and the fact that we do not have the corresponding data for Irish Water, we only replicate models with the explanatory variables group EV2.

- We then re-estimate these three models described above for operating expenditure only (instead of botex). This is because we are not able to accurately identify all IW capital maintenance data at this stage, and therefore consider separate opex and CM models to be more appropriate.

We present these results and discuss limitations below.

3.3. CMA Model Results

Table 3.5 to Table 3.7 present our replication and update of the three CMA models for logarithmic unit cost, linear unit cost and logarithmic aggregate cost models respectively. In each case we have replicated the CMA results exactly for the original full model botex specifications. These results are presented in the first results column of each table.

The second results column shows the results when we re-estimate the full CMA models with opex as the dependent variable instead of botex. As shown in Table 3.5 to Table 3.7, running the models with opex as the dependent variable results in some important changes to the model coefficients. In particular, the coefficient on the regional wage variable now has a negative sign, which is somewhat counter intuitive (it would suggest that higher regional wages result in *lower* expected operating costs). The magnitude of some of the other variables also changes substantially.

Table 3.5
CMA Logarithmic Unit Cost Model – Replication and Adaptation

	(1)	*(2)*
Model / Cost Variable:	*Full Model – Botex (†)*	*Full Model – Opex*
Ln (water delivered per property)	**0.4086**	**0.9389**
	(0.196)	*(0.032)**
Ln (regional wage measure)	**0.6130**	**-0.6699**
	(0.386)	*(0.378)*
Ln (mains length per property)	**0.3996**	**0.0070**
	(0.138)	*(0.97)*
Prop. consumption by metered NHHs	**-0.9979**	**-1.3841**
	(0.114)	*(0.015)**
Proportion of DI from rivers	**0.3333**	**0.4087**
	(0.086)	*(0)****
Proportion of DI from reservoirs	**0.2720**	**0.2202**
	(0.212)	*(0.11)*
Ln (average pumping head)	**0.2487**	**0.2888**
	(0.077)	*(0.01)**
Y = 2009	-0.0226	0.0802
	(0.53)	*(0.031)**
Y = 2010	-0.0079	0.1109
	(0.83)	*(0.015)**
Y = 2011	-0.0180	0.0589
	(0.579)	*(0.127)*
Y = 2012	-0.0138	-0.0118
	(0.446)	*(0.595)*
Constant	-5.7924	-1.5907
	*(0.026)**	*(0.548)*
Adjusted R-Squared	0.4318	0.5486

Source: NERA analysis based on CMA Bristol Final determination
*Statistical significance levels (p-values in parenthesis): * 5%, ** 1%, *** 0.1%.*
(†) Exact replication of the CMA Model with smoothed botex as dependent variable and explanatory variables group EV2.

Table 3.6
CMA Linear Unit Cost Model – Replication and Adaptation

	(1)	*(2)*
Model / Cost Variable:	*Full Model – Botex (†)*	*Full Model – Opex*
Potable water delivered per property	**0.0114**	**0.0420**
	(0.897)	*(0.402)*
Regional wage measure	**0.0031**	**-0.0020**
	(0.466)	*(0.394)*
Mains length per property	**0.0036**	**0.0004**
	(0.083)	*(0.604)*
Prop. consumption by metered NHHs	**-0.1341**	**-0.0987**
	(0.1)	*(0.005)***
Proportion of DI from rivers multiplied by potable water delivered per property	**0.0686**	**0.0524**
	(0.091)	*(0)****
Proportion of DI from reservoirs multiplied by potable water delivered per property	**0.0472**	**0.0244**
	(0.328)	*(0.176)*
Average pumping head multiplied by potable water delivered per property	**0.0004**	**0.0003**
	(0.135)	*(0.023)**
Y = 2009	-0.0013	0.0051
	(0.723)	*(0.017)**
Y = 2010	0.0004	0.0069
	(0.909)	*(0.008)***
Y = 2011	-0.0011	0.0034
	(0.741)	*(0.098)*
Y = 2012	-0.0014	-0.0008
	(0.488)	*(0.549)*
Constant	0.0003	0.0610
	(0.997)	*(0.137)*
Adjusted R-Squared	0.4296	0.5362

Source: NERA analysis based on CMA Bristol Final determination
*Statistical significance levels (p-values in parenthesis): * 5%, ** 1%, *** 0.1%.*
(†) Exact replication of the CMA Model with smoothed botex as dependent variable and explanatory variables group EV2.

Table 3.7
CMA Logarithmic Aggregate Cost Model – Replication and Adaptation

	(1)	*(2)*
Model / Cost Variable:	*Full Model – Botex (†)*	*Full Model – Opex*
Ln (water delivered per property)	**0.4409**	**0.8620**
	(0.182)	(0.06)
Ln (regional wage measure)	**0.5084**	**-0.4213**
	(0.494)	(0.635)
Ln (total mains length)	**1.0144**	**0.9657**
	(0)***	(0)***
Prop. consumption by metered NHHs	**-0.9925**	**-1.3968**
	(0.121)	(0.017)*
Ln (number of properties / total mains)	**0.6006**	**0.9925**
	(0.033)*	(0)***
Proportion of DI from rivers	**0.3312**	**0.4136**
	(0.083)	(0)***
Proportion of DI from reservoirs	**0.2169**	**0.3512**
	(0.384)	(0.138)
Ln (average pumping head)	**0.2538**	**0.2766**
	(0.066)	(0.005)**
Y = 2009	-0.0191	0.0719
	(0.592)	(0.077)
Y = 2010	-0.0044	0.1026
	(0.904)	(0.037)*
Y = 2011	-0.0157	0.0535
	(0.63)	(0.2)
Y = 2012	-0.0159	-0.0067
	(0.399)	(0.778)
Constant	-5.6288	-1.9795
	(0.032)*	(0.48)
Adjusted R-Squared	0.9841	0.9861

Source: NERA analysis based on CMA Bristol Final determination
*Statistical significance levels (p-values in parenthesis): * 5%, ** 1%, *** 0.1%.*
(†) Exact replication of the CMA Model with smoothed botex as dependent variable and explanatory variables group EV2.

3.4. CMA Model Limitations for IRC2 Context

The CMA critique and revised models were developed in the context of a water only company and as a result of Ofwat's comparative benchmarking analysis at PR14. The models have been developed as botex (base operating expenditure) models, which combine both operating expenditure and capital maintenance.

In the context of IW's IRC2 submission, there remain some substantial areas of uncertainty with respect to categorisation of different types of expenditure and in particular IW is not able to accurately identify capital maintenance.[17] As a result we favour separate opex and capital maintenance models.

As discussed above moving from botex models to opex models using the CMA framework results in some large changes in the model results, which in some cases are counter-intuitive. In particular, the coefficient on regional wage becomes negative in all models, which is difficult to explain.

Furthermore, the circumstances facing IW motivate a methodology that is tailored to the specific circumstances facing Irish Water. Specifically, Irish Water is an outlier with respect to scale compared to English and Welsh companies. The CMA models are not constructed to address this factor specifically, which is arguably less important for the E&W companies who are broadly comparable in terms of network length, connections and density.

Running these models with multiple scale variables is problematic due to the high level of multicollinearity between e.g. network length, connections and volumes delivered, which constitutes a risk of bias in the estimated coefficients. For example, we note that applying the coefficients obtained from Model 2 in Table 3.7 to estimate allowed costs would effectively imply assuming a weak, negative relationship between mains length and operating expenditure, which is counter- intuitive.[18]

For these reasons, we develop an alternative set of opex models, which we discuss in Section 4 below.

---

[17] We discuss IW's submission in more detail in the main IRC2 interim review report.

[18] Using the coefficients in Table 3.7, allowed costs would be calculated assuming the relationship: E(ln(opex)) = - 1.98 + *0.966 * ln (mains length) + 0.993 * ln (nr of properties/ mains length)* + [cross-product between the rest of coefficients and explanatory variables] = - 1.98 + *0.966 * ln (mains length) + 0.993 * ln (nr of properties) – 0.993 * ln (mains length)* + […] = - 1.98 + *0.993 * ln (nr of properties) – 0.027 * ln (mains length)* + […], implying an effective coefficient of -0.027 on mains length.

4. Econometric Benchmarking in IRC2

4.1. Data sources and variables used

The data used in our econometric models draws mainly on the dataset Ofwat used for its PR14 models, which provides data on costs and cost drivers in England and Wales for the periods 2009-2013 (for the water service) and 2007-2013 (for the sewerage service).[19] We have added the corresponding water treatment works (WTW) and wastewater treatment works (WWTW) data by combining June Returns and August submissions data.

In order to be able to apply the econometric models to the case of Irish Water (IW), Scottish Water (SW) and Northern Ireland Water (NIW), we have collected the corresponding comparable data from publicly available sources for each of the variables in our models. The data used for IW is based on the IRC2 forward submission. Table 4.1 and Table 4.2 summarise the mapping across jurisdictions between each of the variables used in the IRC2 models.

[19] Ofwat (2014) "Basic Cost Threshold Feeder Model", Appendix A – Water/Wastewater Data Inputs. Available at http://webarchive.nationalarchives.gov.uk/20150624091829/https:/www.ofwat.gov.uk/pricereview/pr14/wholesale/prs_web140404pr14wholesalecostasses

Table 4.1
Data Sources of the Variables Used in the IRC2 Models – Water Service

England and Wales	Scottish Water	Northern Ireland Water	Irish Water
Cost variable			
Wholesale opex, controllable (£m) *(Ofwat PR14 dataset)*	Wholesale opex, controllable (£m) *(Regulatory Accounts, Table M18 W)*	Wholesale opex, controllable (£m) *(Annual Report, Notes to Reg. Accounts, 4a)*	[Total operating expenditure minus Licenses & Levies and VAT (€m)] x 0.5 (assumed water – sewerage split at 50:50)[20] *(BPQ, Tab "6.3 Opex_activities"; VAT figures from FE (2016))*[21]
Cost Drivers			
Water Delivered – potable (Ml/day) *(Ofwat PR14 dataset)*	Net Consumption - including supply pipe losses (Ml/day) *(Annual Returns, Table A2.18)*	Water Delivered (Ml/day) *(NIW Annual Return - Table D)*	Total Water Delivered and Billed + Total Water Taken Unbilled (Ml/day) *(BPQ, Tab "2.1.S&D_W)*
Connected Properties - dom and non-dom (000s) *(Ofwat PR14 dataset)*	Total Connected Properties *(Annual Returns, Table A1.10)*	Connected Properties – excl. void properties *(NIW Annual Return - Table D)*	Household Connected Properties + Non-Household Connected Properties *(BPQ, Tab "5.1.BusAss)*
Mains Length(Km) *(Ofwat PR14 dataset)*	Water Pipes Length (Miles) *(Annual Report and Accounts, p.1)*	Mains Length (Km) *(NIW Annual Return - Table 11)*	Total Length of Mains (Km) *(BPQ, Tab "5.1 BusAss")*
Real Regional Wage (£/h) *(Ofwat PR14 dataset, based on ONS)*	Real Regional Wage (£/h) *(NERA calculation based on Ofwat procedure and ONS data. See further details below)*	Real Regional Wage (£/h) *(NERA calculation based on Ofwat procedure and ONS data. See further details below)*	Real Regional Wage (£/h) *(NERA calculation based on Ofwat procedure and CSO data. See further details below)*
Water Treatment Works (Nr) *(June Returns for 2009-11, then rolled forward for 2012-13)*	Water Treatment Works (Nr) *(Annual Report and Accounts, p.1)*	Water Treatment Works (Nr) *(NIW Annual Return - Table 12)*	Water Treatment Plants (Nr) *(IRC2 Opex Submission, p.3)*
Average Pumping Head (m.hd) *(Ofwat PR14 dataset)*	Avg. Pumping Head – Resources and Treatment + Avg. Pumping Head – Water Distribution (m.hd) *(Annual Returns, Tables E4.14 and E6.25)*	Average Pumping Head (m.hd) *(NIW Annual Return - Table 12)*	Energy Proxy (m.hd) / Distribution Input (Ml/day) *(IW Opex Benchmarking 2016, Exhibit 9; BPQ, Tab "2.1.S&D_W")*

Source: NERA analysis of various sources (see cells in table)

[20] We understand that in the UK, water companies' VAT costs are recoverable and, for this reason, reported costs exclude VAT. Therefore, in order to put all opex levels into comparable terms, we also exclude VAT from IW's submitted opex.

[21] Frontier Economics (20 April 2016) "Irish Water Opex Benchmarking" – a confidential report for Irish Water.

Table 4.2
Data Sources of the Variables Used in the IRC2 Models - Sewerage Service

England and Wales	Scottish Water	Northern Ireland Water	Irish Water
Cost variable			
Wholesale opex, controllable (£m) *(Ofwat PR14 dataset)*	Wholesale opex, controllable (£m) *(Regulatory Accounts, Table M18 WW)*	Wholesale opex, controllable (£m) *(Annual Report, Notes to Reg. Accounts, 4a)*	[Total operating expenditure minus Licenses & Levies and VAT (€m)] x 0.5 (assumed water – sewerage split at 50:50)[22] *(BPQ, Tab "6.3 Opex_activities"; VAT figures from FE (2016))*[23]
Cost Drivers			
Load Received by Sewage Treatment Works (kg BOD/ day) *(Ofwat PR14 dataset)*	Total Load (kg BOD/ day) *(Annual Returns, Table E7.31)*	Load Entering Sewerage Syst. (to BOD/yr) Converted into kg BOD/day by multiplying by the SW's average (2009-13) ratio between Total Load (kg BOD/day – *Table E7.31*) and Total Load (tonnes BOD/yr – *Table A3.21*). *(NIW Annual Return - Table 15)*	Load received at sewage treatment works (Kg BOD/day)[24] *(BPQ, Tab "2.2.S&D_WW")*
Connected properties (000s) *(Ofwat PR14 dataset)*	Connected Properties (000s) *(Annual Returns, Table A1.21)*	Connected Properties (000s) *(NIW Annual Return - Table E)*	Household Connected Properties + Non-Household Connected Properties (000s) *(BPQ, Tab "5.1.BusAss")*
Sewers length (Km) *(Ofwat PR14 dataset)*	Sewer Pipe Length (Miles) *(Annual Report and Accounts, p.1)*	Sewers Length (Km) *(NIW Annual Return - Table 16)*	Length of Wastewater Pipelines (Km) *(IRC2 Opex Subm., p.3)*
Real Regional Wage (£/h) *(Ofwat PR14 dataset)*	Real Regional Wage (£/h) *(NERA calculation based on Ofwat procedure and ONS data. See further details below)*	Real Regional Wage (£/h) *(NERA calculation based on Ofwat procedure and ONS data. See further details below)*	Real Regional Wage (£/h) *(NERA calculation based on Ofwat procedure and CSO data. See further details below)*
Wastewater Treatment Works (Nr) *(June Returns for 2007-11, August Subm. for 2012-13)*	Wastewater Treatment Works (Nr) *(Annual Report and Accounts, p.1)*	Sewage Treatment Works (Nr) *(NIW Annual Return - Table 15)*	Wastewater Treatment Plants (Nr) *(IRC2 Opex Submission, p.3)*

Source: NERA analysis of various sources (see cells in table)

22 See footnote 20.

23 Frontier Economics (20 April 2016) "Irish Water Opex Benchmarking" – a confidential report for Irish Water.

24 The units indicated in IW's BPQ are Kg BOD/year, instead of Kg BOD/day. However, given its scale, we consider that the value only makes sense if it is expressed in Kg BOD/day, and therefore assume a mistake in the units given. Figure A.7 in Appendix A shows that, under this assumption, IW's data point is consistent with its UK comparators.

Constructing the regional wage indices

Ofwat's calculates regional wage indices drawing on the Office for National Statistic's *Annual Survey of Hours and Earnings* (ASHE), which provides average regional hourly earnings by occupation category, excluding overtime pay.[25]

Ofwat weights two types of occupations at the 2-digit SOC level to form its regional index for water and sewerage utilities:

- **40 per cent:** *"Specialist: 21 - Science, Research, Engineering and Technology Professionals"*;
- **60 per cent:** *"Skilled: 53 - Skilled Construction and Building Trades"*

In the calculation of the regional wage for IW, SW and NIW we have followed the procedure adopted by Ofwat (as described above), as closely as permitted by data availability.

For Scotland and Northern Ireland we use the same ASHE data as for England and Wales, available from the ONS and the DETI respectively.[26]

For Ireland we use the Central Statistic's Office's *Earnings and Labour Costs*, which provides average hourly earnings excluding irregular earnings by sector.[27] The occupational breakdown in the CSO data is different to that in the ONS ASHE data. We have applied the following weightings, which provide an approximate mapping to Ofwat's procedure:

- **40 per cent:** *"M - Professional, scientific and technical activities"*;
- **60 per cent:** *"F – Construction"*.

For the summary stats presented in Table 4.3, we have converted this into GBP pounds per hour using the OECD's Purchasing Power Parities (PPP) for private consumption (we discuss the use of PPP for conversion between currencies below).

4.2. Summary statistics of key driver variables

In Table 4.3 we present summary statistics of Irish Water's operations alongside UK comparator statistics.[28] The bars on the right side show IW in green, and the England and Wales average in light blue. As shown, IW largely falls within the comparator range and we would therefore expect statistical models based on these data to provide a reliable estimate of IW's expected cost levels. There are however some exceptions to this which it is worth noting:

---

[25] Ofwat (2014) "Cost Assessment – Advanced Econometric Models", Annex 3. Final report submitted by CEPA.

[26] Northern Ireland's Department of Enterprise, Trade and Investment (2014) *"ASHE for NI by industry, occupation, age, public/private sector and skill level, 2014 (provisional) and 2013 (revised)"*. We use 2013 data for consistency with Ofwat's dataset.

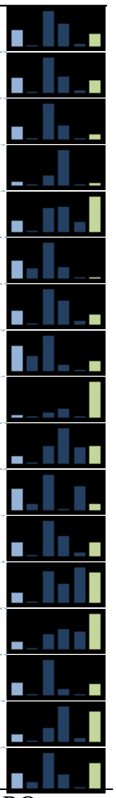
[27] We use 2013 data for consistency with Ofwat's dataset.

[28] Note that this chart shows comparable data for UK WASCs only – we present full summary stats by driver variable for all companies including the water only companies (WOCs) in Appendix A.

- IW's water mains length is more than 30 per cent higher than the closest comparator in the UK (Scottish Water);
- IW's water property density (connected properties / mains length) is lower than all UK comparators, although relatively close to Northern Ireland Water's density;
- IW's number of water treatment works is more than three times higher than the closest UK comparator (Scottish Water);
- IW's leakage as a proportion of distribution input is nine percentage points higher than the closest comparator in the UK (Scottish Water); [29]
- The regional wage we have constructed for Ireland appears to be above the UK average, although well within the broader pack of comparators.

Table 4.3
IW and UK Comparator Summary Statistics

	E&W Ave	E&W Min	E&W Max	SW	NIW	Irish Water
Water Connections (000)	2,054	593	3,678	2,601	761	1,706
Water Population Served (000)	4,619	1,261	9,162	5,002	1,814	4,063
WW Connections (000)	2,422	712	5,640	2,479	618	1,342
WW Population Served (000)	5,650	1,624	14,541	49,993	1,024	3,374
Mains Length (km)	28,325	11,610	46,576	48,135	26,499	62,951
Water Property Density (Nr/km)	70	51	118	54	29	27
Sewer Pipes (km)	32,736	9,235	68,591	49,993	15,090	25,000
WW Property Density (Nr/km)	74	60	87	50	41	54
Water Treatment Works (Nr)	88	39	138	252	24	918
WW Treatment Works (Nr)	638	350	1,123	1,865	1,024	1,102
Average Pumping Head (m.hd)	128	78	176	58	140	78
Water Delivered (Ml/d)	942	271	2,053	1,291	441	945
Usage (Ml/property per day)	0.45	0.41	0.56	0.50	0.58	0.55
Leakage/DI (proportion)	0.21	0.15	0.27	0.31	0.29	0.43
Sewage Load (Kg BOD/day)	380,695	100,938	915,978	222,744	72,506	328,396
Area Served (km^2)	21,409	7,162	35,550	80,239	14,130	69,825
Regional Wage (£/h)	15.59	14.26	18.62	15.35	13.27	17.06



Source: NERA analysis of Ofwat PR14 dataset, August Submissions, Regulatory Accounts and IW's BPQ response

Table 4.4 and Table 4.5 show (for the water and sewerage service respectively) the variables for which we have data available for all companies, and which consider potentially relevant explanatory variables for econometric cost models. The tables present the correlation between the explanatory variables.

[29] We do not include this variable in any of the IRC2 econometric models.

For both services, the three first explanatory variables are highly correlated. This is not surprising, as they are all variables that measure the scale of activity undertaken. This high degree of multicollinearity can lead to instability in model coefficients for correlated variables, and we consider alternative model specifications to address this concern below.

Table 4.4
Correlations between Explanatory Variables - Water Service

	Water delivered	Connections	Mains length	Regional wage	WTW	Avg. pumping head
Water delivered	1.000					
Connections	0.980	1.000				
Mains length	0.866	0.931	1.000			
Regional wage	0.150	0.039	-0.178	1.000		
WTW	0.641	0.676	0.758	0.040	1.000	
Avg. pumping hd.	-0.316	-0.301	-0.183	-0.023	-0.074	1.000

Source: NERA analysis of Ofwat's PR14 dataset and other E&W regulatory submissions.

Table 4.5
Correlations between Explanatory Variables - Sewerage Service

	Water delivered	Connections	Sewers length	Regional wage	WWTW
Load (kg BOD5)	1.000				
Connections	0.997	1.000			
Sewers length	0.976	0.981	1.000		
Regional wage	0.546	0.565	0.496	1.000	
WWTW	0.046	0.063	0.221	-0.355	1.000

Source: NERA analysis of Ofwat's PR14 dataset and other E&W regulatory submissions.

4.3. Methodology

As at IRC1 our approach is to draw on the data available from English and Welsh water and sewerage companies which have been subject to incentive based regulation for over a decade and represent a reasonable benchmark for efficient cost performance which IW should achieve over time. The results of the econometric models estimated drawing on E&W companies are used to predict what the efficient level of costs should be for each company, including for Scottish Water and Northern Ireland Water.

4.3.1. The Composite Scale Variable

Given the concerns around multicollinearity of the scale variables described in the context of the CMA models in Section 3 and as shown by the summary statistics for the current dataset in Section 4.2, we have developed models which draw on a composite scale variable (CSV). The CSV provides a single driver that combines various different scale variables into a single scale variable and is a way of mitigating multicollinearity concerns in subsequent regressions. This method has been used by a number of utility regulators, including Ofgem and NIAUR.[30]

In creating the CSV, we follow Ofgem's methodology at RIIO-ED1, by combining the different scale variables using a weighted geometric mean. For example, for the water service the CSV takes the form:[31]

$$CSV = (Water\ Delivered)^{w1}\ x\ (Connected\ Properties)^{w2}\ x\ (Mains\ Length)^{w3},$$

where w1, w2 and w3 are weightings on each of the scale variables. We discuss sensitivity to the weights used, and an approach to determining "optimal" weights, below.

We have estimated the "optimal" weights following a procedure used by Ofgem (2014). This procedure is based on the following regression:

$$Ln\ opex = Constant + b_1\ x\ Std\ (ln\ Water) + b_2\ x\ Std\ (ln\ Properties) + b_3\ x\ Std\ (ln\ Length),$$

which is intended to measure the relative impact of each variable on costs.[32] Note that the log scale variables have been standardised in order to avoid a driver with a large average having an undue effect on the calculation of the weights.[33]

The resulting weights are of the form $w_n = b_n / (b_1 + b_2 + b_3)$. Table 4.6 summarizes the results we have obtained following this methodology. Set 1 corresponds to the weights obtained when including the three variables in the regression. Set 2 excludes water delivered and load for the water and the sewerage services respectively.

Table 4.6
Estimated CSV Weights Following Ofgem (2014) Methodology

Water			Sewerage		
	Set 1	Set 2		Set 1	Set 2
Water delivered	0.40	-	Load	-1.08	-
Connections	0.43	0.90	Connections	1.26	0.30
Mains Length	0.17	0.10	Sewers Length	0.83	0.70

Source: NERA Analysis.

30 Ofgem (2014) *"RIIO-ED1: Final determinations for the slow-track electricity distribution companies"*, Appendix 5
Utility Regulator (2014) *"Water and Sewerage Services Price Control 2015-21"*, Final Determination - Annex Q
Deloitte (2016) *"Relative efficiency of Northern Ireland Gas Distribution Network"*, Annex 4

31 The same applies to the sewerage service, using the variables: treated load, connected properties, and sewers length

32 We note that this regression can entail multicollinearity and omitted variable problems, which can lead to some bias in the coefficients. However, it is useful as a first-stage approximation of the relative importance of each driver in explaining the variation in costs.

33 *Standardising* a variable consists in subtracting the average of the variable from each observation and dividing by the standard deviation. Then, the new standardised variable has an average of 0 and a standard deviation of 1. By standardising all the explanatory variables in a regression, we can estimate the effect of each variable on costs independently of the size of their average.

Broadly, the weights can be interpreted as the relative importance of each of the scale factors in the CSV, and in all cases the weights sum to 1. For example, for Set 1 of water weights, the CSV would effectively be comprised of 40 per cent water delivered, 43 per cent connections and 17 per cent mains length.

In Set 1 for sewerage, load has a negative coefficient. This is counter-intuitive and may result from multicollinearity, as discussed above. For this reason, we use Set 2 as the starting point for our models.

The Utility Regulator in Northern Ireland adopted a similar approach to comparative benchmarking at PC15.[34] UREGNI included a CSV model with optimal weights calculated following the same procedure as above. Table 4.7 shows UREGNI's optimal weights (Set 1), along with the other two sets of weights (Set 2 and Set 3) that NIAUR includes in its range of alternative models.

The set of variables used is slightly different from that used in our models. In particular, UREGNI uses distribution input instead of water delivered, and population served instead of connected properties. However, for the water service, its results are comparable to those we have obtained, where mains length is statistically attributed the lowest weight among the three variables.

For the sewerage service, UREGNI also uses population served in preference to connections. UREGNI also obtains a positive coefficient for load and therefore retains load in its optimal weightings (Set 1) and in its sensitivities (Set 2 and Set 3). Noting these differences to our optimal weighting results described in Table 4.6, we include these weights in the suite of models that we run for Irish Water, as discussed in Section 4.4.

Table 4.7
CSV Weights Used in NIAUR's PC15 Alternative Efficiency Models

Water				Sewerage			
	Set 1	Set 2	Set 3		Set 1	Set 2	Set 3
Distribution Input	0.66	0.40	0.33	Load	0.26	0.40	0.33
Population	0.26	0.40	0.33	Population	0.44	0.40	0.33
Mains Length	0.08	0.20	0.33	Sewers Length	0.30	0.20	0.33

Source: NERA summary based on Sections 5, 6 and 8 of Utility Regulator (2014) "Water and Sewerage Services Price Control 2015-21", Final Determination - Annex Q.

34 Utility Regulator (2014) *"Water and Sewerage Services Price Control 2015-21"*, Final Determination - Annex Q

4.3.2. Regional Wage Adjustment

Regional differences in wage levels may drive differences in company costs for companies operating in different regions. To the extent that these wage differentials are outside of the companies' control, this should be accounted for in comparative efficiency analysis. There are two possible approaches to making such an adjustment:

- Off-model adjustments as adopted by Ofgem in RIIO-ED1[35] and RIIO-GD1,[36] by UR in PC15,[37] and by the CC in the 2014 NIE decision.[38]
- Including regional wage as an explanatory variable in the econometric models, as implemented by Ofwat in PR14.[39]

We favour the first of these approaches which involves scaling up or down in order to improve comparability across companies before conducting cost benchmarking. This approach has been commonly adopted by regulators, and has the advantage that it avoids any risks of statistical bias in the estimation of the regional wage effect.[40]

Step 1: Calculate regional wage indices.
The first step in this procedure is to calculate an index that provides, for each year and company, the ratio between the average wage in England and Wales in that year and the regional wage in that company. This implies that if, for example, a company's regional wage equals 95% of the E&W average, that company's index is 1.053 (1.0/0.95).

Step 2: Determine the proportion of opex which relates to internal labour costs
The index calculated in Step 1 is then applied to the share of operating expenditure that corresponds to labour costs.[41] Using the latest data available that provides a breakdown of E&W companies' operating expenditure by functional area, we have estimated that the average share of labour costs in the E&W water sector is approximately 30% for the water service, and 40% for the sewerage service.[42] In our definition of labour costs, we include both in-house employment costs and hired and contracted services. Ideally, we would only make the regional adjustment to costs that must be incurred at a local level, excluding hired services that need not be co-located with the network and hence could be outsourced outside the company's region.[43] We do not have sufficiently granular data to isolate this effect, and

---

35 Ofgem (2014): *RIIO-ED1 Final determinations for the slow-track electricity distribution companies - Business plan expenditure assessment*, 28 November 2014, para. 4.1

36 Ofgem (2012): *RIIO-GD1: Final Proposals – Supporting document – Cost efficiency,* 17 December 2012, para. 2.1

37 UR (2014): *Water & Sewerage Services Price Control 2015-21 Final Determination - Annex P*, December 2014, para. 5.1.2.

38 CC (2014): *Northern Ireland Electricity Ltd Price Determination, Final Determination*, 26 March 2014, para 8.67.

39 CEPA (2014): *Cost Assessment – Advanced Econometric Models*, 20 March 2014, page 56

40 We also experimented with including regional wage as an explanatory variable in the econometric models but found that the coefficient was not robust to model specification and varied not just in magnitude but in sign.

41 By using the formula: [Adjustment factor] = 1 – [Share of labour costs] + [Share of labour costs] x [Wage Differential Index].

42 Ofwat June Returns data, E&W average over years 2008-10.

43 See e.g. Ofgem in RIIO-GD1, where only a percentage of work is considered to be required to be done locally.

note that in the context of Irish Water the local labour market is likely to be the most relevant market for all labour service – both internal and outsourced.

Step 3: Run econometric models on wage-adjusted company operating costs.
Running the econometric models on the adjusted operating expenditure data will provide, for each company, its predicted costs as if the company was placed in a hypothetical region with average wage levels.

Step 4: Re-adjust modelled costs to account for the regional wage differential.
As a final step, we re-adjust the predicted costs to reflect the regional wage level as calculated in Step 1 in order for the results to be comparable with the companies' actual costs. This is done by dividing the predicted costs by the same pre-modelling adjustment factor.

For IW, we calculate real labour costs are around 8% higher than the England and Wales, which we allow for in our efficiency assessment. Our real labour adjustment for IW is higher than for all comparator WaSCs, with the exception of Thames Water. (See Appendix A.)

Table 4.8 and Table 4.9 show, for each Water and Sewerage Company (WaSC), how the costs predicted by our preferred econometric model, including the off-modelling adjustment for regional wages, compare to observed costs (IRC2 submitted costs in the case of IW). Our real labour adjustment index for IW results in an adjustment to Irish Water's model costs of €4.1m for the water service, and €5.5m for the sewerage service.

Table 4.8
Regional Wage Adjustments by WaSC – Water Service

Company	Model Costs (average wage levels)	Adjustment	Predicted Costs (own regional wage levels)	Observed costs	Efficiency Gap
Wessex Water	44.9	-0.5	44.4	40.4	-9%
Southern Water	73.7	1.3	74.9	53.4	-29%
South West Water	58.5	-0.6	57.9	56.7	-2%
Welsh Water	105.6	-3.4	102.2	124.3	22%
Anglian Water	153.4	-1.0	152.5	137.1	-10%
Northumbrian Water	141.7	-1.8	139.9	142.2	2%
Yorkshire Water	157.7	-2.0	155.7	145.8	-6%
United Utilities	215.6	-4.2	211.3	184.2	-13%
Severn Trent Water	232.0	-4.2	227.8	264.8	16%
Thames Water	258.8	12.0	270.8	320.8	18%
NI Water	69.4	-3.8	65.6	84.0	28%
Scottish Water	199.0	-1.9	197.1	160.1	-19%
Irish Water	**186.0**	**4.0**	**190.1**	**350.3**	**84%**

Source: NERA Analysis. All values are in € million, 2015 prices.

Table 4.9
Regional Wage Adjustments by WaSC – Sewerage Service

Company	Model Costs (average wage levels)	Adjustment	Predicted Costs (own regional wage levels)	Observed costs	Efficiency Gap
Wessex Water	96.9	-0.2	96.6	72.8	-25%
South West Water	68.6	-0.3	68.2	80.7	18%
Northumbrian Water	94.5	-2.5	92.0	85.5	-7%
Welsh Water	120.0	-3.7	116.3	111.8	-4%
Southern Water	118.1	3.4	121.5	137.9	13%
Yorkshire Water	161.1	-1.2	159.8	166.2	4%
United Utilities	206.1	-3.3	202.9	215.1	6%
Severn Trent Water	270.9	-3.5	267.4	226.3	-15%
Anglian Water	228.1	1.5	229.6	232.6	1%
Thames Water	262.9	14.3	277.3	324.5	17%
NI Water	93.6	-5.4	88.1	117.2	33%
Scottish Water	262.9	-0.9	261.9	131.0	-50%
Irish Water	**176.8**	**6.1**	**182.9**	**350.3**	**92%**

Source: NERA Analysis. All values are in € million, 2015 prices.

4.3.3. Post-Modelling Adjustments

To put the model results on a comparable basis to the data submitted by Irish Water in its IRC2 submission, we make a number of post-modelling adjustments:

- We express all costs in 2015 prices. We have converted all UK company costs into 2015 prices using annual averages of the RPI CHAW data provided by ONS, and all Irish Water costs using Irish HICP series provided by IW. As a result, we have increased UK modelled and observed costs by 3.4%, and IW modelled costs by 0.4%.
- We convert the 2015 price base UK cost data into euros using the OECD's Purchasing Power Parities (PPP) for private consumption. PPPs are the rates of currency conversion that eliminate the differences in price levels between countries.[44] We use the 2012-14 average, which has a value of 1.25 €/£.[45]
- We adjust Irish Water's modelled cost to reflect its retail activities. The PR14 models and database are of wholesale costs. To provide an estimate of the modelled costs for an integrated utility providing both wholesale and retail services we uprate IW's modelled costs by a factor of 1.25. This value consists on the average ratio of total to wholesale-only operating expenditure across all the WaSCs in England and Wales, based on analysis of recent regulatory accounts.
- We exclude VAT from IW's submitted costs. In the UK water companies' VAT costs are recoverable and, for this reason, reported costs exclude VAT. Therefore, in our comparisons with IW's modelled costs, IW's submitted costs also exclude VAT.

  In IW's IRC2 submissions, the only VAT included as a specific line item is "irrecoverable VAT" which we understand relates to VAT on items procured at the Ervia group level. In the absence of more reliable data, and as a conservative approach, we implement the values used by Frontier Economics in their external benchmarking report, submitted by IW as part of the formal Q&A process with the CER.[46]

4.4. Results

We considered various combinations of explanatory variables in our models, including CSV variables with various weightings, water and waste water treatment works, and average pumping head

All of our models are in logarithmic terms, with total controllable operating expenditure as the explanatory variable. Our estimation method is a pooled OLS regression in common with Ofwat, Ofgem and CMA modelling approach at recent reviews. Table 4.10 and Table 4.11 show the resulting coefficients (and their p-values in parenthesis) of each of the various models and sensitivities we have estimated for each service.

---

44 See http://www.oecd.org/std/purchasingpowerparitiespppsdata.htm

45 This is not substantially different from the Bank of England exchange rate, which has a 2013-April 2016 average of 1.27 €/£

46 Frontier Economics (20 April 2016) *"Irish Water Opex Benchmarking"* – a confidential report for Irish Water.

Our approach to model selection has been to start from the long-list of potential variables included by Ofwat at PR14 and subsequently by the CMA. We include the CSV variable as the main cost driver in all of our models.

We then start with a full model specification, based on the data made available to us from IW and consider refined models with fewer explanatory variables where the fuller models include factors which are not statistically significant. The approach and results for the water models are described in Section 4.4.1 and the approach and results for the waste water models are described in Section 4.4.2.

4.4.1. Model results – water service

For the water service, the full specification that we take as our starting point is to include the number of water treatment works and the average pumping head alongside the CSV.[47] In contrast to the models developed for IRC1, we do not include leakage in our models. Leakage is (at least over time) endogenous, i.e. it is within the company's control. High leakage may result in high opex, but this may also reflect inefficiency if the level of leakage is above the economic level of leakage (ELL).[48] We would expect this to be the case in Ireland where IW has inherited a system with relatively high levels of leakage compared to UK water utilities. This may represent a legitimate reason for IW's costs to be somewhat higher today, but also represents an opportunity to lower operating costs over time.

The number of water treatment works would be expected to raise costs, all other factors held constant. As shown in the results for Model 1, we find a negative coefficient on water treatment works – which we also found to be the case at IRC1. It is possible that the number of treatment works acts as a proxy variable for density. Density could result in lower levels of costs given distances will be shorter, or indeed increase in costs where the operator in a dense network faces higher congestion costs. However, we exclude water treatment works in our preferred specification (Model 3), as it is not clear that we are modelling a causal relationship rather than picking up noise from any omitted variables (however, all models presented in Table 4.10 are included in the modelled ranges presented in **Error! Reference source not found.**).

Average pumping head reflects topography and the pumping requirements to deliver water to connections across the network. A higher pumping head would be expected to increase energy costs and therefore raise opex. In the case of IW, the inclusion of this variable would decrease allowed costs, since IW's average pumping head is lower than the E&W average. We do not find a significant statistical relationship in our models for average pumping head

---

47 Average pumping head is intended to measure how much each megalitre of water is pumped through the process, from abstraction to supply, and is defined as $\Sigma_i (l_i * WP_i)/ (V_p + V_g)$, where l_i is the annual mean lift (in meters) at site *i*, WP_i is the volume of water pumped at site *i*, and $(V_p + V_g)$ is the total volume of water that enters supply. We include average pumping head in preference to total pumping head as total pumping head is largely driven by volume of water, which is accounted for in the CSV scale variable. The CMA models also included average pumping head.

48 The CMA also does not include leakage in its models, arguing that *"Including leakage as an explanatory variable in a model might increase the extent to which the model 'explains' differences in costs between companies but may reduce the extent to which the model captures differences between companies in their efficiency of providing water services to customers"* − CMA (2015) Bristol Water Final Determination, Appendix 4.2, par. 136.

when water treatment works are not included in the model (Model 2). We therefore omit average pumping head from our preferred specification (Model 3)

We include year dummies, taking 2013 as the reference year, following the approach taken by the CMA, although these are in some cases not significant. The year dummies should account for any specific shock in a particular year that may have changed the cost environment and would also pick up any general trend in costs over time.

Finally, the main cost driver, the composite scale variable, includes mains length, properties connected and water delivered.

Model 4 to Model 6 present sensitivities around the weightings used for the CSV variable, for our preferred specification (Model 3). We consider the following weights for water delivered, connections and mains length, respectively:

- Model 4 (40:40:20). As shown in Table 4.7, NIAUR uses these weights in one of its alternative efficiency models for PC15. The rationale behind it is the fact that water costs present a much higher correlation with volumes and population than with network length.[49]
- Model 5 (0:75:25). Since the confidence in IW's estimates of water delivered levels is relatively low compared to the other two scale variables, we include a model that does not take into account this variable. For the same reasons outlined above, we give mains length a lower weight with respect to connected properties.
- Model 6: (33:33:33). For completeness, and consistent with UREGNI's models, we include a CSV model that gives equal weights to all variables.

[49] Utility Regulator (2014) *"Water and Sewerage Services Price Control 2015-21"*, Final Determination - Annex Q

Table 4.10
IRC2 Econometric Model Results - Water

	Optimal CSV Weights			***Sensitivity of CSV Weights***		
	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*
CSV weights						
Water delivered	*40*	*40*	*40*	*40*	*-*	*33*
Connections	*43*	*43*	*43*	*40*	*75*	*33*
Mains	*17*	*17*	*17*	*20*	*25*	*33*
ln_csv	**1.155**	**0.999**	**0.992**	**0.992**	**0.980**	**1.002**
	*(0.000)****	*(0.000)****	*(0.000)****	*(0.000)****	*(0.000)****	*(0.000)****
ln_avg_pumping_head	**0.212**	**0.191**				
	*(0.049)**	*(0.167)*				
ln_WTW	**-0.173**					
	*(0.014)**					
Y = 2009	0.062	0.054	0.048	0.048	0.072	0.051
	*(0.004)***	*(0.011)**	*(0.019)**	*(0.019)**	*(0.001)***	*(0.014)**
Y = 2010	0.094	0.089	0.088	0.087	0.110	0.090
	*(0.000)****	*(0.001)****	*(0.001)****	*(0.001)****	*(0.000)****	*(0.001)****
Y = 2011	0.045	0.048	0.050	0.050	0.075	0.053
	*(0.011)**	*(0.009)***	*(0.008)***	*(0.000)**	*(0.000)****	*(0.005)***
Y = 2012	0.007	0.009	0.014	0.015	0.031	0.017
	(0.671)	*(0.574)*	*(0.383)*	*(0.380)*	*(0.072)*	*(0.307)*
Constant	-4.474	-7.868	-2.955	-3.054	-3.394	-3.452
	*(0.000)****	*(0.000)****	*(0.000)****	*(0.000)****	*(0.000)****	*(0.000)****
Adjusted R-Squared	0.979	0.971	0.968	0.968	0.966	0.967
Nr. of observations	90	90	90	90	90	90

Source: NERA analysis.
*Statistical significance levels (p-values in parenthesis): * 5%, ** 1%, *** 0.1%.*

We use the coefficients in Table 4.10 to predict the level of opex that each company would be expected to undertake according to the observable characteristics of its water service When we run these predictions we also predict expected cost ranges for Scottish Water, Northern Ireland Water and Irish Water, whose data is not included in running the econometric models.

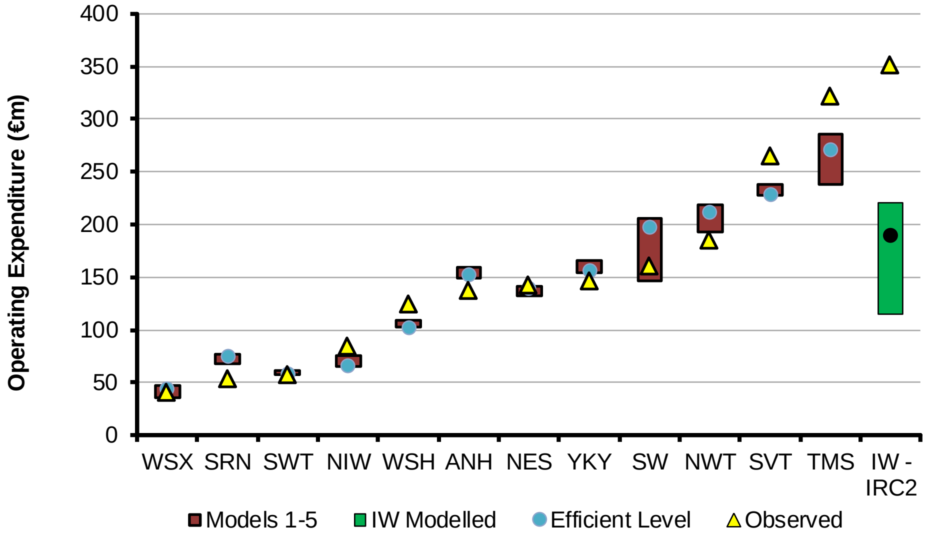
Figure 4.1 presents the modelled cost ranges for the water service. The bars represent the modelled costs as predicted by our models, while the triangles represent actual costs.[50] The round blue markers present the results from our preferred specification.

The modelled cost range for IW is large compared to the other UK companies. This is because IW is currently an outlier with respect to some of the cost drivers; notably with respect to mains length and water treatment works.

50 For the comparator companies, the actual costs are outturn costs from 2013. For Irish Water the outturn costs are the average over the two years of the IRC2 submission. All values are expressed in 2015 prices.

IW's proposed controllable (and net of VAT for comparability) operating expenditure of €350m is €129m (58%) above the upper end of our modelled opex (i.e. under the model most favourable to IW).[51] IW's proposed controllable (and net of VAT for comparability) opex is €160m (84%) higher than under our preferred model specification.

Figure 4.1
IRC2 Opex Model - Water Modelled Ranges



Source: NERA analysis of Ofwat's PR14 dataset, other UK regulatory submissions and IW's IRC2 submissions.

Note: Models estimated using E&W data (2009-13). All UK modelled and observed costs presented in this chart correspond to 2013; all IW costs correspond to the 2017-18 average. All costs are in 2015 prices.

4.4.2. Model results – waste water

For the waste water service, the full specification that we take as our starting point is to include the number of waste water treatment works alongside the CSV.[52] This is our preferred model (Model 2 in Table 4.11).

The number of waste water treatment works would be expected to affect costs in much the same way as the number of water treatment works would affect costs as described above. In general, we would expect more WWTW to raise costs, which is what we find in the coefficient in our models. We therefore retain WWTW alongside the CSV in our preferred specification. As with the water models described above, we include year dummies for each year of data included in the models.

---

51 See footnote 20.

52 Average pumping head is intended to measure how much each megalitre of water is pumped through the process, from abstraction to supply, and is defined as $\Sigma_i (l_i * WP_i)/ (V_p + V_g)$, where l_i is the annual mean lift (in meters) at site *i*, WP_i is the volume of water pumped at site *i*, and $(V_p + V_g)$ is the total volume of water that enters supply. We include average pumping head in preference to total pumping head as total pumping head is largely driven by volume of water, which is accounted for in the CSV scale variable. The CMA models also included average pumping head.

Finally, the main cost driver, the composite scale variable, includes pipe length, properties connected and sewage load.

Model 3 to Model 7 present sensitivities around the weightings used for the CSV variable. We consider the following weights for sewage load, connections and sewers length, respectively:

- Model 3 (0:50:50), Model 4 (0:75:25) and Model 5 (0:25:75). Keeping the weight on load to zero for the reasons outlined above, we explore different combinations of weights on connections and length.
- Model 6 (26:44:30). As shown in Table 4.7, NIAUR estimates these optimal weightings in one of its alternative models for PC15, using a similar procedure to the one described in Section 4.3.1. Noting the differences with our own set of estimated optimal weightings, we consider appropriate to test the sensitivity of our results to the use of NIAUR's weightings.
- Model 7 (33:33:33). For further robustness and completeness, and in consistency with NIAUR's models, we include a model that gives equal weights to all variables.

Table 4.11
IRC2 Econometric Model Results - Sewerage

	Optimal CSV Weights		***Sensitivity of CSV Weights***				
	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*	*Model 7*
CSV weights							
Load	*-*	*-*	*-*	*-*	*-*	*26*	*33*
Connections	*30*	*30*	*50*	*75*	*25*	*44*	*33*
Sewer length	*70*	*70*	*50*	*25*	*75*	*30*	*33*
ln_csv	**0.747**	**0.726**	**0.728**	**0.731**	**0.725**	**0.724**	**0.730**
	*(0.000)****	*(0.000)****	*(0.000)****	*(0.000)****	*(0.000)****	*(0.000)****	*(0.000)****
ln_WWTW		**0.203**	**0.226**	**0.255**	**0.197**	**0.248**	**0.244**
		*(0.033)**	*(0.014)**	*(0.005)***	*(0.040)**	*(0.007)***	*(0.009)***
Y = 2007	-0.001	-0.001	0.002	0.005	-0.001	-0.007	-0.010
	(0.984)	*(0.987)*	*(0.972)*	*(0.919)*	*(0.977)*	*(0.884)*	*(0.827)*
Y = 2008	0.015	0.015	0.017	0.019	0.014	0.011	0.009
	(0.720)	*(0.718)*	*(0.683)*	*(0.638)*	*(0.728)*	*(0.781)*	*(0.825)*
Y = 2009	0.030	0.031	0.032	0.033	0.030	0.031	0.030
	(0.476)	*(0.471)*	*(0.451)*	*(0.427)*	*(0.476)*	*(0.467)*	*(0.480)*
Y = 2010	0.068	0.068	0.070	0.071	0.068	0.071	0.071
	(0.128)	*(0.125)*	*(0.118)*	*(0.109)*	*(0.127)*	*(0.113)*	*(0.114)*
Y = 2011	0.004	0.004	0.005	0.006	0.004	0.005	0.005
	(0.878)	*(0.844)*	*(0.817)*	*(0.783)*	*(0.851)*	*(0.812)*	*(0.823)*
Y = 2012	-0.035	-0.036	-0.035	-0.034	-0.036	-0.035	-0.035
	(0.138)	*(0.137)*	*(0.141)*	*(0.148)*	*(0.135)*	*(0.152)*	*(0.153)*
Constant	-3.870	-9.838	-5.715	-6.712	-4.710	-6.081	-5.828
	*(0.007)***	*(0.002)***	*(0.000)****	*(0.000)****	*(0.001)***	*(0.000)****	*(0.000)****
Adjusted R-Squared	0.883	0.913	0.920	0.927	0.911	0.922	0.919

Source: NERA analysis.

*Statistical significance levels (p-values in parenthesis): * 5%, ** 1%, *** 0.1%.*

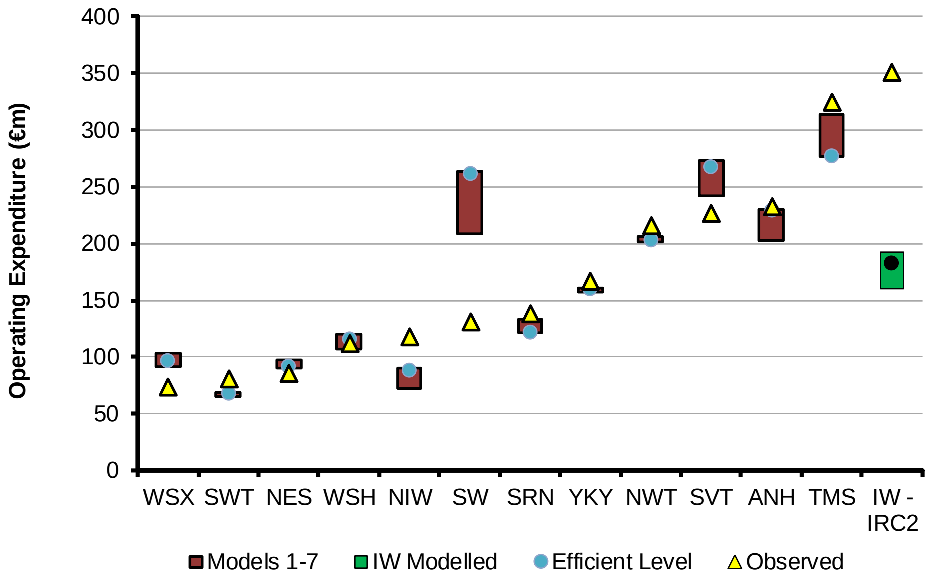
We use the coefficients in Table 4.11 to predict the level of opex that each company would be expected to undertake according to the observable characteristics of its waste water service. When we run these predictions we also predict expected cost ranges for Scottish Water, Northern Ireland Water and Irish Water, whose data is not included in running the econometric models.

Figure 4.2 presents the modelled cost ranges for the waste water service. The bars represent the modelled costs as "predicted" by our models, while the triangles represent actual costs.[53] The round blue markers present the results from our preferred specification.

Unlike water, the modelled cost range for IW is not particularly large compared to the other UK companies. This is because IW is not an outlier with respect to the scale cost drivers; it is an outlier with respect to waste water treatment works but this does not have a material impact on costs.

IW's proposed controllable (and net of VAT for comparability) opex expenditure of €350m is €158m (82%) above the upper end of our modelled opex (i.e. under the model most favourable to IW).[54] IW's proposed controllable (and net of VAT for comparability) opex expenditure is €167m (92%) higher than under our preferred model specification.

Figure 4.2
IRC2 Opex Model - Sewerage Modelled Ranges



Source: NERA analysis of Ofwat's PR14 dataset, other UK regulatory submissions and IW's IRC2 submissions.

Note: Models estimated using E&W data (2007-13). All UK modelled and observed costs presented in this chart correspond to 2013; all IW costs correspond to the 2017-18 average. All costs are in 2015 prices.

53 For the comparator companies, the actual costs are outturn costs from 2013. For Irish Water the outturn costs are the average over the two years of the IRC2 submission. All values are expressed in 2015 prices.

54 See footnote 20.

5. IW's Comparative Analysis

As part of the formal Q&A process with the CER, IW has submitted an external benchmarking report which updates the analysis undertaken by NERA at IRC1.[55]

The study, undertaken by Frontier Economics, states that its approach has been "based as far as possible on NERA's methodology". However, there are some substantial changes made to the NERA model, and the FE approach does not appear to reflect developments from recent comparative benchmarking exercises undertaken by UK regulators. We summarise the key elements of FE's methodology and their results below.

5.1. IW's comparative benchmarking methodology and results

Frontier have carried out a cross-section OLS approach, which consists of running a separate regression using data for each of the years between 2014 and 2018. FE makes a number of pre-modelling adjustments to put costs for all companies on comparable terms, namely excluding retail costs, subtracting uncontrollable opex items, subtracting VAT and an off-model adjustment for regional wage differences.

In its water model, IW includes log mains length, log network density leakage and total pumping head as cost drivers.[56] For the wastewater service it includes log sewer length, log network density, log load and the number of wastewater treatment works. The study does not present results from any other than the preferred specifications described above.

FE find that IW's gap to the efficiency frontier for water was 17% in 2014, growing to 35% in 2018. For the sewerage service the efficiency gap was 26% in 2014, growing to 32% in 2018

5.2. NERA review of IW comparative benchmarking

Our review of the Frontier Report on benchmarking IW's opex expenditure proposals gives rise to the following high-level comments.

- **Only one model is presented, with no detailed explanation for why this model is preferred or the process that resulted in this model specification**. FE uses mains length as its principal cost driver. It concludes that the results from using connections as an alternative driver to network length "were not substantially different from our preferred model"
- **The models are run including Irish Water in the modelling dataset**. Our approach has been to model the mature utilities of England and Wales which have been subject to incentive based regulation for some time. We then project modelled results for Scottish Water, Northern Ireland Water and Irish Water. Including IW in the models will by the

55 Frontier Economics (20 April 2016) *"Irish Water Opex Benchmarking"* – a confidential report for Irish Water

56 By total pumping head we mean average pumping head times total water input

nature of statistical modelling move the efficiency frontier closer to IW, which we do not consider appropriate given IW's early stage of development. Our approach is consistent with the approach adopted by utility regulators in considering the efficiency of NIW and SW at an early stage of development.

- **FE uses a forecast dataset from 2014 – 2018** based on projections of growth in wholesale opex. This is inconsistent with regulators' approaches. Ofwat (PR14) and CMA (Bristol Water) estimates models drawing on historical data series. Ofgem has partially drawn on forecast data (e.g. Ofgem's approach to benchmarking at RIIO-ED1) but places at least equal emphasis on models estimated using historical data. There is no clear reason to draw on forecast data, which undermines the robustness of estimated models (as forecast expenditure is subject to forecast error).
- **FE uses ordinary least squares (OLS) as opposed to panel or pooled data approaches**. In the past, Ofwat used cross-sectional data but all GB regulators (Ofwat, Ofgem, CMA) draw on panel or pooled data approaches. Specifically, we use a pooled model across the years for which there are data which improves the robustness of the models relative to FE's OLS models.[57]
- **FE makes a number of pre-modelling adjustments to IW's submitted costs, but does not provide any of the modelling calculations or detailed methodology for these adjustments**:
  - Adjustment for regional wages, of c. €60m off c. €750m. This is significantly higher than the adjustment of €20.6m (€8m for the water service, and €12.6m for the sewerage service) equivalent to the adjustment that NERA has estimated in Section 4.3.2.[58]

    FE's adjustment is based on differences in indices of average weekly earnings for electricity, gas and water supply sector from Ireland's Central Statistics Office and the UK's Office for National Statistics. As noted by Ofwat in PR14, the problem with using water industry indices is that companies have the ability to directly influence that data, which entails endogeneity problems.[59] We use indices for relevant occupations as opposed to the indices for the industry which are affected by water companies. We also use hourly wages as opposed to weekly wage comparisons to control for differences in hours worked in Ireland relative to GB regions.

    Furthermore, FE has not set out its assumed difference in real wages based on its selected indices (to compare to our estimate of real wages 8% higher than GB average).
  - Reduction for VAT – c. €50m. We have no comparable VAT data from IW in its IRC2 submission, where the only VAT as a specific line item is *"irrecoverable VAT"*

57 For example, FE includes leakage in its models for each year from 2014 to 2018 and the coefficient changes substantially over this period.

58 For the purpose of comparison with FE's adjustment, we present our adjustment applied to IW's *submitted* costs. Therefore, the values differ from those presented in Section 4.3.2 (€4.1m for the water service, and €5.5m for the sewerage service), which correspond to the equivalent adjustment applied to IW's *modelled* costs.

59 CEPA (2014): *Cost Assessment – Advanced Econometric Models*, 20 March 2014, page 56

which we understand relates to VAT on items procured at the Ervia group level. FE does not provide any details on the basis on which this adjustment has been calculated. In the absence of more reliable data, and as a conservative approach, we have applied the same VAT adjustment in our models.

- Reduction for retail of c. €50m. This appears to be based on the sum of expenditure lines which can be identified as being related to retail-specific activities in the IW submission, such as billing and customer service. This is a relatively small proportion of total opex (6%) and is smaller than the proportion we assume (20% of total opex).[60]

For these reasons we do not consider the exercise IW has undertaken to be an improvement on the models we have developed in Section 2 to Section 4 above.

6. Conclusions

We have used a range of econometric benchmarking models to compare IW's proposed expenditure to expenditure levels of UK companies. Based on this analysis, we conclude that IW's proposed operating expenditure is high compared to a benchmark level of efficient expenditure formed based on UK water and sewerage companies.

In developing our models, we have taken into account Irish Water specific factors, notably around the inclusion of network length as a cost driver within a composite scale variable, as well as adjusting for higher real wages in Ireland relative to UK. By contrast, we have not taken into account factors that advantage IW in the comparison, e.g. the lower levels of environmental compliance in Ireland which reduces IW costs relative to UK peers. In addition, we have measured IW's costs relative to the predicted (or average) costs. By contrast, regulators in UK typically measure efficiency relative to upper quartile company costs.

We have developed a number of models acknowledging the difficulty with specifying a definitive cost function and model for water companies. In all cases, the model results need to be interpreted with caution as the difference between IW actual and expected costs may reflect statistical error as well as relative efficiency e.g. the model specification may not fully take into account IW operating characteristics. However, we consider that a reasonable interpretation of the evidence is that IW's costs are at least 70% higher than the long run efficient level, and our preferred modelling approach shows that its costs are around 100% higher than the long-run efficient level.

60 This value consists on the average ratio of retail-only to total operating expenditure across all the WaSCs in England and Wales, based on analysis of recent regulatory accounts. Under this assumption, we apply an equivalent post-modelling adjustment of 125% (total to wholesale-only opex) on IW's modelled wholesale costs, as described in Section 4.3.3.

Appendix A. Cost Driver Summary Statistics

This appendix illustrates all the cost drivers data for England and Wales used in the estimation of our econometric models, along with the corresponding data points for Irish Water, Scottish Water and Northern Ireland Water.

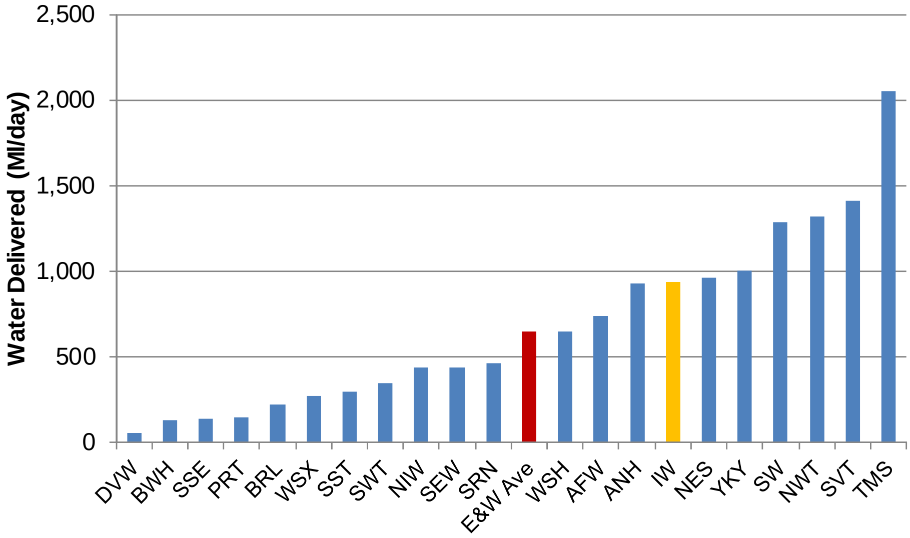
All the UK data corresponds to 2013, except from the Irish Water data, which refers to the data submitted for the IRC2 period (2017-18). Further details on the data sources can be found in Table 4.1 and Table 4.2.

A.1. Water Service

This section presents the cost drivers included in our econometric models for the water service − either directly, or indirectly through the use of the Composite Scale Variable (CSV) described in Section 4.3.1. It also includes the regional wage variable used in the calculation of the indices for the off-modelling regional wage adjustment described in Section 4.3.2.

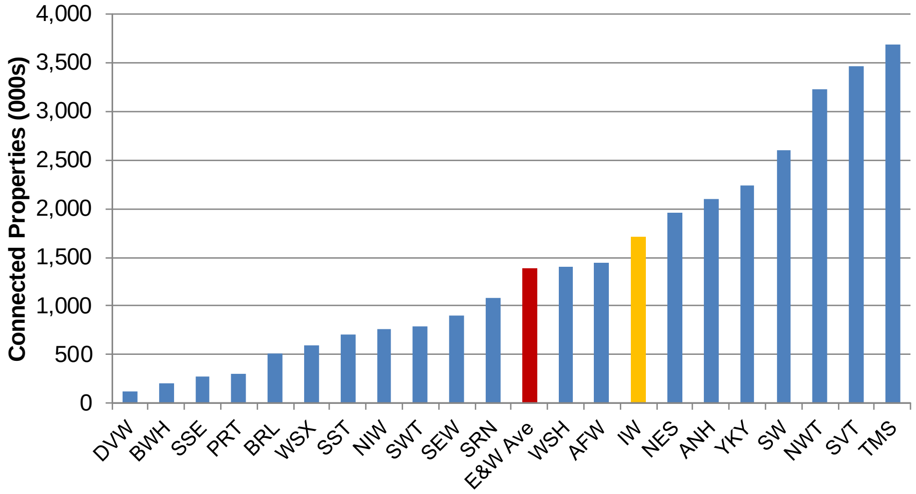
We present the data used for all the companies in England and Wales − including the Water only Companies (WoCs) −, along with the corresponding data for Irish Water, Scottish Water and Northern Ireland Water.

Figure A.1
UK Comparator Statistics − Water Delivered



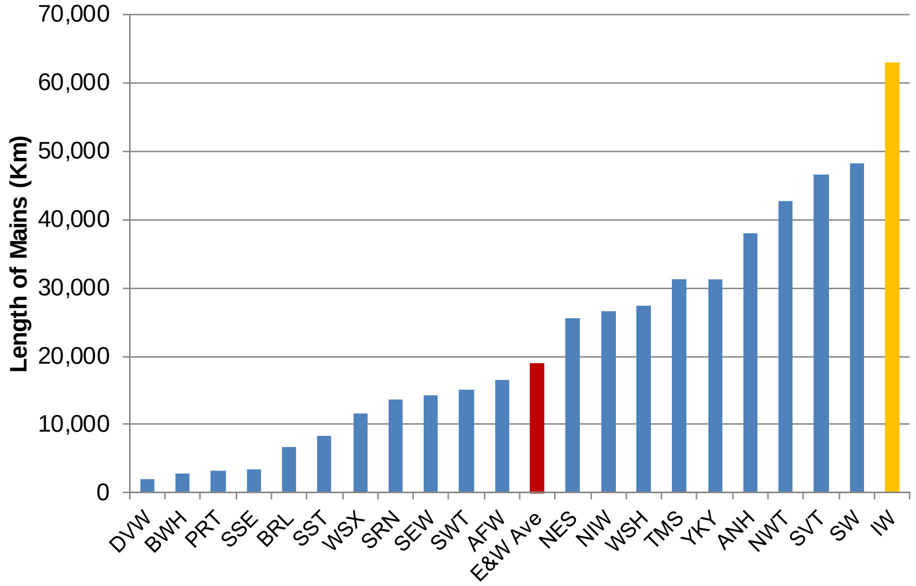
NERA analysis of Ofwat PR14 econometric models, company accounts and IW's IRC2 submissions.

Figure A.2
UK Comparator Statistics – Connected Properties



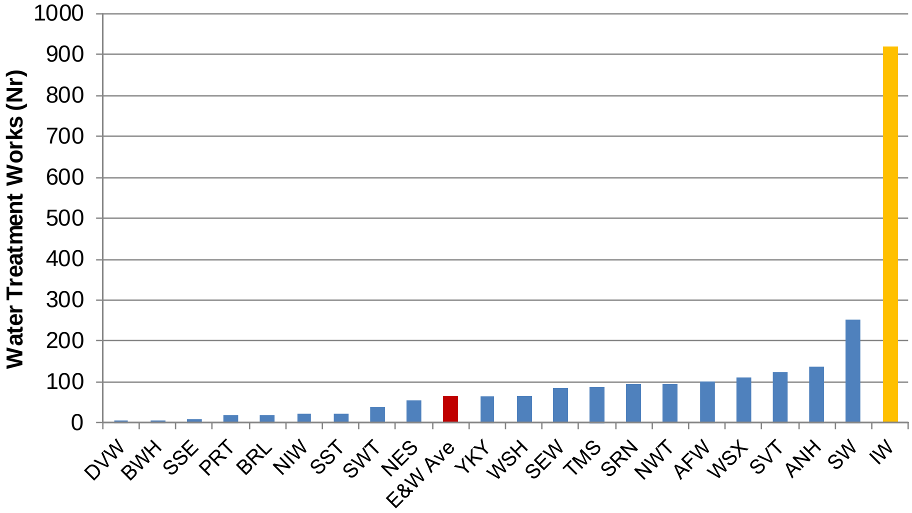
NERA analysis of Ofwat PR14 econometric models, company accounts and IW's IRC2 submissions.

Figure A.3
UK Comparator Statistics – Length of Mains



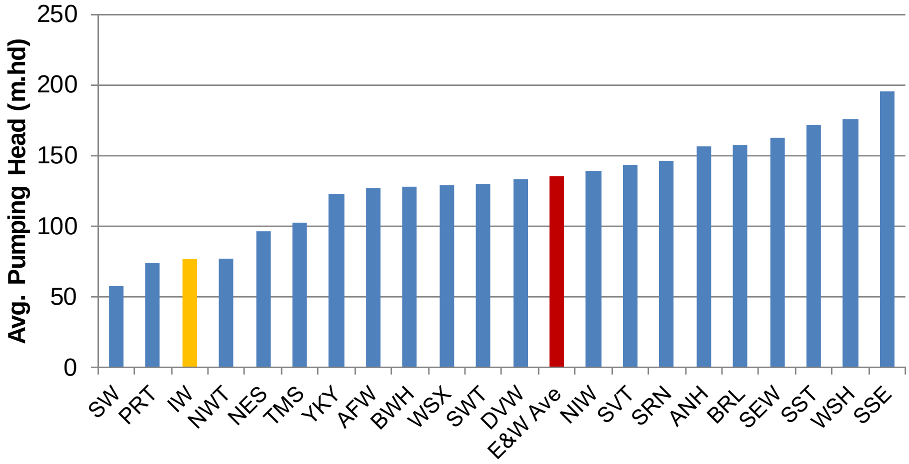
NERA analysis of Ofwat PR14 econometric models, company accounts and IW's IRC2 submissions.

Figure A.4
UK Comparator Statistics – Water Treatment Works



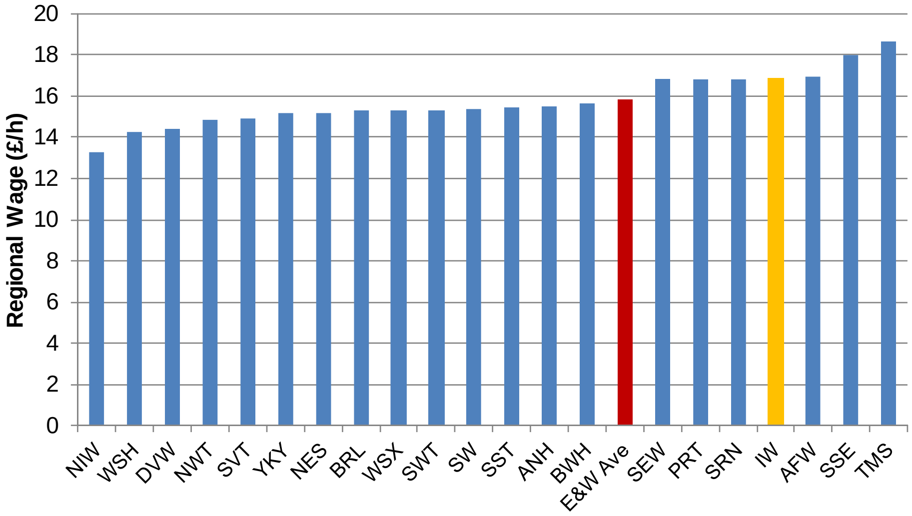
NERA analysis of Ofwat June Returns, company accounts and IW's IRC2 submissions.

Figure A.5
UK Comparator Statistics – Average Pumping Head



NERA analysis of Ofwat PR14 econometric models, company accounts and IW's IRC2 submissions.

Figure A.6
UK Comparator Statistics − Regional Wage



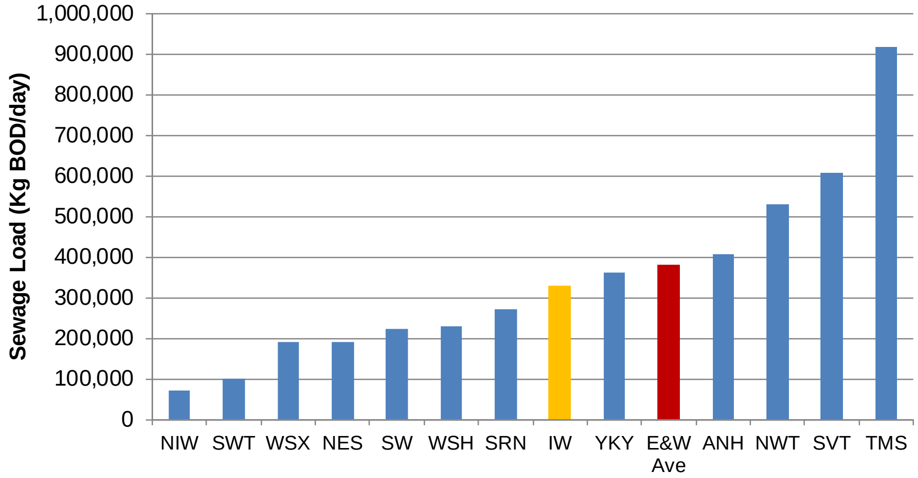
NERA analysis of Ofwat PR14 econometric models, company accounts and IW's IRC2 submissions.

A.2. Sewerage Service

This section presents the cost drivers included in our econometric models for the sewerage service − either directly, or indirectly through the use of the Composite Scale Variable (CSV) described in Section 4.3.1. It also includes the regional wage variable used in the calculation of the indices for the off-modelling regional wage adjustment described in Section 4.3.2.

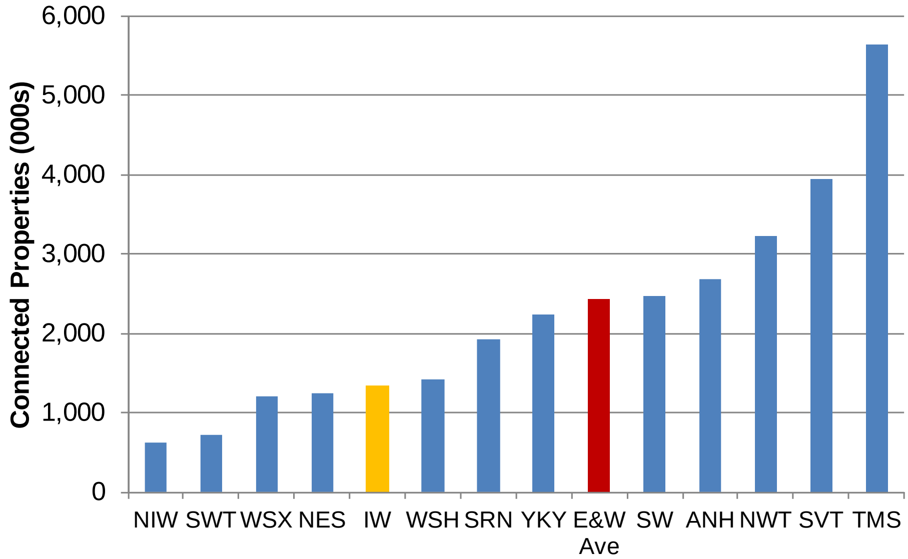
We present the data used for all the Water and Sewerage Companies (WaSCs) in England and Wales, along with the corresponding data for Irish Water, Scottish Water and Northern Ireland Water.

Figure A.7
UK Comparator Statistics – Sewage Load



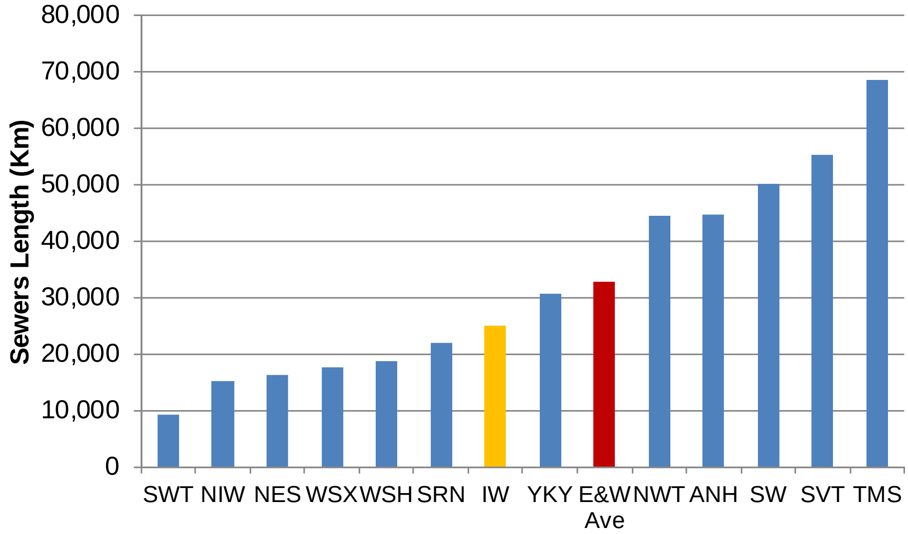
NERA analysis of Ofwat PR14 econometric models, company accounts and IW's IRC2 submissions.

Figure A.8
UK Comparator Statistics – Connected Properties



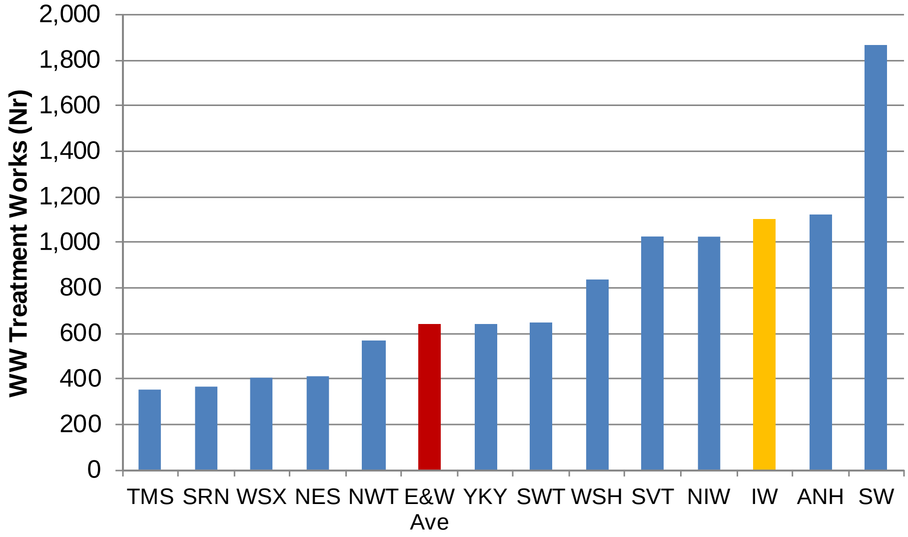
NERA analysis of Ofwat PR14 econometric models, company accounts and IW's IRC2 submissions.

Figure A.9
UK Comparator Statistics − Sewers Length



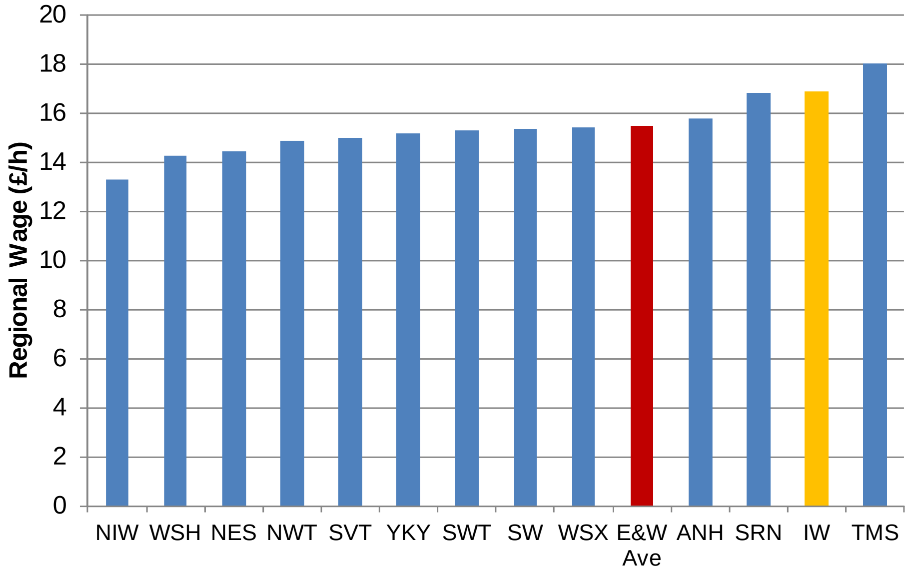
NERA analysis of Ofwat PR14 econometric models, company accounts and IW's IRC2 submissions.

Figure A.10
UK Comparator Statistics − Wastewater Treatment Works



NERA analysis of Ofwat June Returns, August Submissions, company accounts and IW's IRC2 submissions.

Figure A.11
UK Comparator Statistics − Regional Wage



NERA analysis of Ofwat PR14 econometric models, company accounts and IW's IRC2 submissions.

Report qualifications/assumptions and limiting conditions

This report is for the exclusive use of the NERA Economic Consulting client named herein. This report is not intended for general circulation or publication, nor is it to be reproduced, quoted or distributed for any purpose without the prior written permission of NERA Economic Consulting. There are no third party beneficiaries with respect to this report, and NERA Economic Consulting does not accept any liability to any third party.

Information furnished by others, upon which all or portions of this report are based, is believed to be reliable but has not been independently verified, unless otherwise expressly indicated. Public information and industry and statistical data are from sources we deem to be reliable; however, we make no representation as to the accuracy or completeness of such information. The findings contained in this report may contain predictions based on current data and historical trends. Any such predictions are subject to inherent risks and uncertainties. NERA Economic Consulting accepts no responsibility for actual results or future events.

The opinions expressed in this report are valid only for the purpose stated herein and as of the date of this report. No obligation is assumed to revise this report to reflect changes, events or conditions, which occur subsequent to the date hereof.

All decisions in connection with the implementation or use of advice or recommendations contained in this report are the sole responsibility of the client. This report does not represent investment advice nor does it provide an opinion regarding the fairness of any transaction to any and all parties.

NERA
ECONOMIC CONSULTING

NERA Economic Consulting
Marble Arch House, 66 Seymour Street
London W1H 5BT
United Kingdom
Tel: 44 20 7659 8500 Fax: 44 20 7659 8501
www.nera.com

NERA UK Limited, registered in England and Wales, No 3974527
Registered Office: Marble Arch House, 66 Seymour Street, London W1H 5BT