Water & Sewerage Services Price Control 2015-21

Annex Q – Alternative Efficiency Modelling

July 2014

Contents page

Water and Sewerage Services Price Control 2015-21

Annex Q – Alternative Efficiency Modelling

1.0 Introduction 1

1.1. Problems 1
1.2. Solution 1
1.3. Model types 2
1.4. Assumptions 3

2.0 Unit Costs – Properties 4

2.1. Water properties 4
2.2. Sewage properties 5
2.3. Unit cost efficiency gap 5

3.0 Unit Costs – Volumes 7

3.1. Water volumes 7
3.2. Sewage volumes 8
3.3. Unit cost efficiency gap 9

4.0 Population Topex Regression 10

4.1. Water and sewage population 10
4.2. Population topex model efficiency gap 11

5.0 CSV Topex Model A 12

5.1. Water CSV model A – equal weights 12
5.2. Sewage CSV model A – equal weights 13
5.3. CSV topex model efficiency gap – equal weights 14

6.0 CSV Topex Model B 15

6.1. Water CSV model B – changed weights 15
6.2. Sewage CSV model B – changed weights 16
6.3. CSV topex model efficiency gap – different weights 17

7.0 Summary of Findings..18

7.1. Comparing approaches..18

1.0 Introduction

1.1. Problems

1.1.1 The efficiency gap has been assessed used the corrected ordinary least squares (COLS) models. The findings suggest a gap estimate of 23%. This means that NI Water would have to reduce costs by this amount to be a frontier company.

1.1.2 Within the business plan, the company has cited a number of concerns with COLS. These include:

a) Models have become outdated and less robust over time;

b) The method used to allow for real opex changes may introduce bias as each companies costs do not move uniformly;

c) The frontier companies may now be different due to changing opex; and

d) Inclusion of PPP costs increases uncertainty of the models.

1.1.3 Since June Return data stopped being published, benchmarking has become harder to do. The COLS models are a cross-sectional comparison at a particular point in time. Unfortunately, the present approach means comparing updated NI Water data with older England and Wales information.

1.1.4 The Utility Regulator (UR) recognises the issue and has tried to address it to some extent within the current approach. Analyses of regulatory accounts help to provide changes in company and industry opex. These changes are reflected by adjusting average and frontier positions.

1.1.5 However, it is accepted that an issue still exists. The models do not include updated explanatory data. Some bias may also be introduced by simply amending predicted cost by the industry change.

1.1.6 To correct for these issues, the UR has completed further efficiency modelling.

1.2. Solution

1.2.1 The PR14 August data submission has provided updated (2012-13) cost and physical data for the England and Wales companies. The submission is not as detailed as June Returns. They do however allow for high-level comparison between companies.

1.2.2 In order to test the robustness of the COLS findings, the UR has developed some total opex (topex) models. These can be used to triangulate around an efficiency figure given that no one method is perfect.

1.2.3 To help improve model precision, the latest data has been pooled with historic June Returns. The UR has used figures from 2008-09 to 2012-13 for the ten WaSC’s. This gives 50 observations against which NI Water can be compared.

1.2.4 The purpose of this annex is to explain the ‘pooled COLS’ models, their rationale and findings. By considering updated models, the UR can test if the COLS problems are adversely affecting the efficiency gap analysis.

1.2.5 The benefit of developing new models is that they provide an independent check on the COLS findings. This can help verify the results or identify areas where concern may arise.

1.3. Model types

1.3.1 A number of potential models are investigated. The options reflect the type of analyses that could be undertaken to support the COLS findings and aid target setting.

1.3.2 In this paper, five alternative options are presented, though any number of variations exists. The options include:

a) **Properties Unit Cost** – This compares NI Water’s cost per property against the England and Wales industry average and frontier;

b) **Volume of Water/Wastewater Unit Cost** – The second method again represents a simple unit cost comparison. This method differs only in the use of volumes rather than properties;

c) **Population Topex Model** – A log regression using total opex as the dependent variable and population served as the single explanatory factor;

d) **CSV Topex Model A** – This option regresses log opex against a composite scale variable. The variable differs between the water and sewage models but includes factors considered key cost drivers. Each is given an equal weighting; and

e) **CSV Topex Model B** – A similar log regression to Model A but explores the use of different weights.

1.3.3 The models were chosen for a variety of reasons. The unit cost method has appeal in that it is a relatively simple approach, which is easy to grasp.

1.3.4 The denominators (properties and volumes) are similar to those previously used by Ofwat to compare companies’ opex costs in their annual *Water and Sewerage Service Unit Costs and Relative Efficiency* reports.

1.3.5 The population regression estimates the relationship between costs and the population served. It is anticipated the two variables will be closely correlated.

1.3.6 CSV Model A uses a composite of various variables. Combining these variables has the benefit of including their impact without introducing a correlation problem that may exist with a number of independent variables.

1.3.7 CSV Model B uses the same approach but varies the weights allocated to each input.

1.4. Assumptions

1.4.1 The models have been developed using five years of pooled data, from 2008-09 to 2012-13. Nominal cost is inflated to current prices. The adopted approach provides more observations. This should help improve model specifications and accuracy.

1.4.2 In the interest of simplicity and comparability with the COLS models, a number of assumptions have been made. These include:

- Total opex for all models excludes business activity costs and other removed costs e.g. rates, service charges, third party costs etc;
- NI Water observations are excluded from the calculation of the unit cost industry average and the topex regressions;
- Special factors and atypical costs are the same as those used in the COLS analysis;
- Residual adjustments remain the same as COLS (10% for water models and 20% for sewerage); and
- Frontier companies remain as Yorkshire Water and Wessex Water for the water and sewage areas.

1.4.3 At this stage, the assumptions are for the benefit for simplicity. They may not however reflect what a more robust approach might look like.

2.0 Unit Costs – Properties

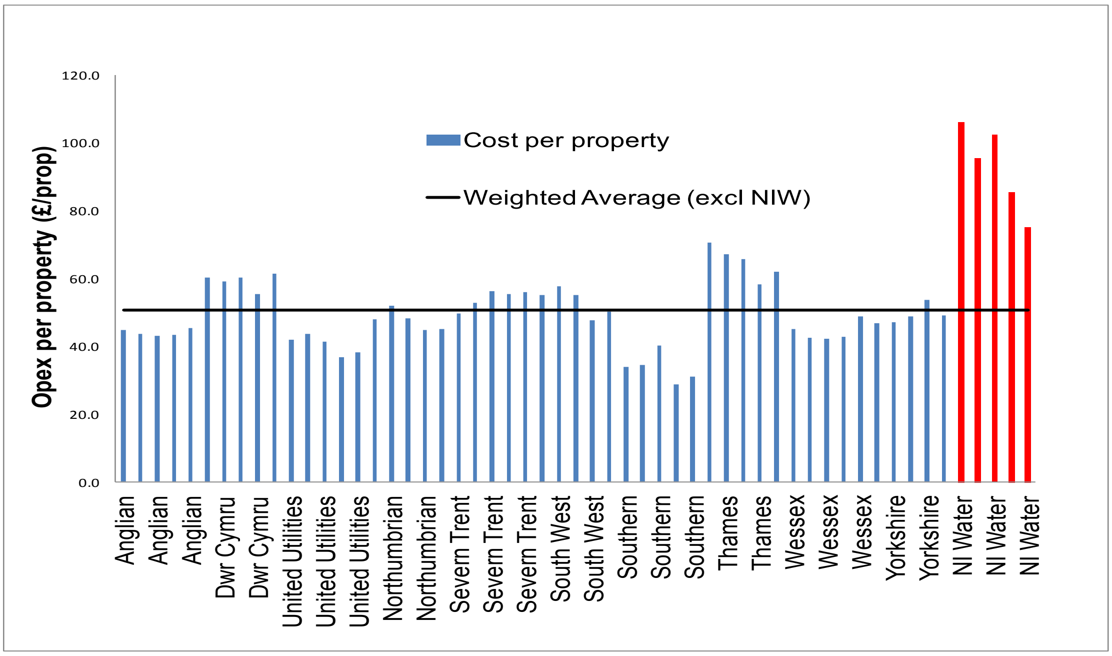
2.1. Water properties

2.1.1 The simplest method of benchmarking is unit costs. This involves dividing costs by a variable that drives expenditure. In this case, connected properties is the chosen variable as billed property data is restricted.

Table 2.1 – Water service cost per connected property

Water Service:	**Total opex (excluding business activities)**
Data:	June Returns and PR14 August submission
Unit cost model:	The unit cost reflects the weighted industry opex cost per billed property. Comparison is made of functional expenditure less service charges and business activities against predicted costs (connected properties multiplied by the weighted average industry unit cost).
£/property	Weighted average industry unit cost: £50.64[1]
Number of observations	50

Figure 2.1 – Water service cost per connected property



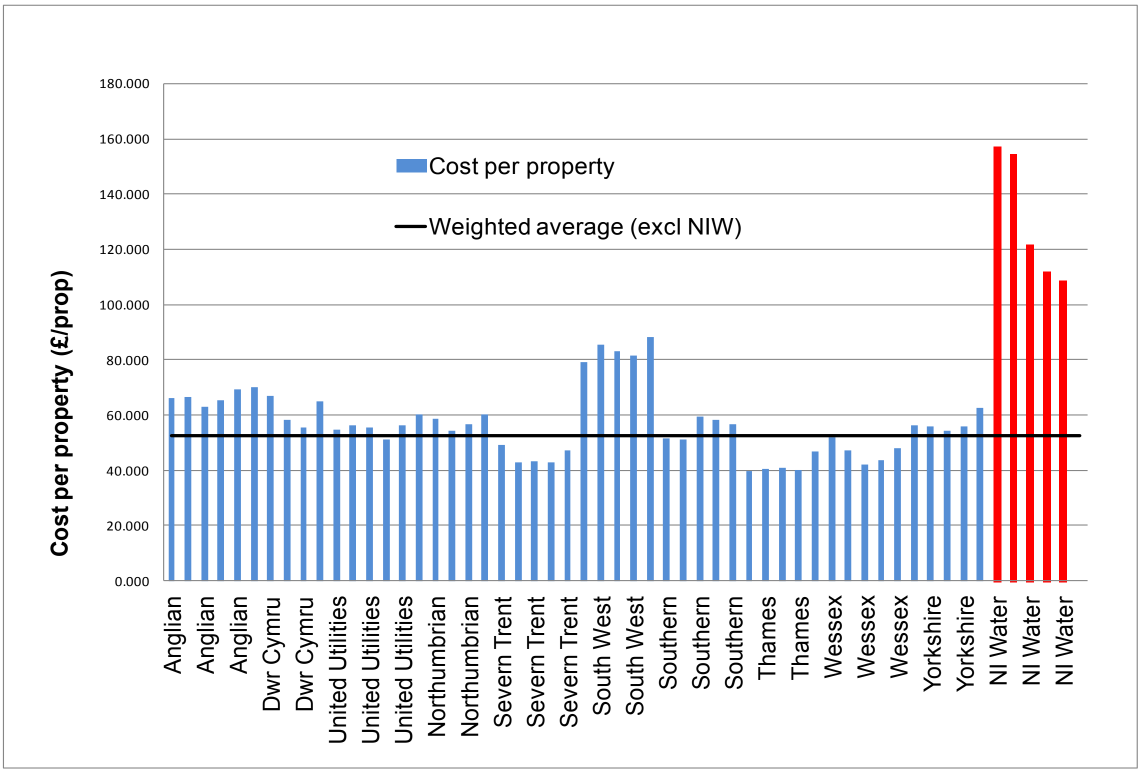
[1] All financial figures in this report are given in 2012-13 prices unless otherwise stated.

2.1.2 Comparisons show NI Water improvement over time. They also indicate much higher spending levels than comparators.

2.2. Sewage properties

2.2.1 Billed properties are used as the denominator for the sewage model. The story is similar to water, though the gap is greater.

Figure 2.2 – Sewage service cost per billed property



2.3. Unit cost efficiency gap

2.3.1 Calculation of the efficiency gap follows the same process as the COLS models. NI Water predicted cost is found by multiplying the industry average cost by its own property data.

Table 2.2 – Predicted cost calculations for water and sewage

Special Factor Claim	Water 2012-13	Sewage 2012-13
NI Water properties (000's)	818.00	618.38
Industry average unit costs (£/prop)	50.64	52.64
NI Water predicted cost (£m)	41.43	32.55

2.3.2 Residual adjustments are then applied alongside a frontier correction. This gives an efficiency gap as follows:

Table 2.3 – Properties unit cost efficiency gap for 2012-13

Efficiency Gap Calculation (2012-13 prices)					
	Category	**Process Rule**	**Water (£m)**	**Sewerage (£m)**	**Total (£m)**
A	NI Water actual cost		61.52	67.34	128.86
B	Less Adjustments		-3.01	-3.29	-6.30
C	**Modelled Cost**	**A – B**	**58.51**	**64.05**	**122.56**
D	Predicted Cost (average)		41.43	32.55	73.98
E	Difference	C – D	17.08	31.50	48.58
F	Adjustment Factor (%)		10%	20%	
G	Residual Adjustment	E * F	1.71	6.30	8.01
H	**New Predicted Costs**	**D + G**	**43.13**	**38.85**	**81.99**
I	Frontier Adjustment (%)		-2.51%	-7.22%	
J	**Frontier Predicted Costs**	**H * (1 + I)**	**42.05**	**36.05**	**78.10**
K	Efficiency Gap (to average)	C – H	15.38	25.20	40.57
L	**Efficiency Gap % (to average)**	**K / C**	**26.28%**	**39.34%**	**33.11%**
M	Efficiency Gap (to frontier)	C – J	16.46	28.00	44.46
N	**Efficiency Gap % (to frontier)**	**M / C**	**28.13%**	**43.72%**	**36.28%**

Figures may not sum due to rounding.

2.3.3 The findings suggest inefficiency levels quite a bit larger than the COLS analysis. Not much faith is placed in these results. While special factors are included to the same extent as the COLS models, simple unit costs do not properly account for (dis)economies of scale.

2.3.4 The selection of the explanatory variable is also crucial. The next models highlight the difference in the efficiency gap by replacing properties with volumes as the denominator.

3.0 Unit Costs – Volumes

3.1. Water volumes

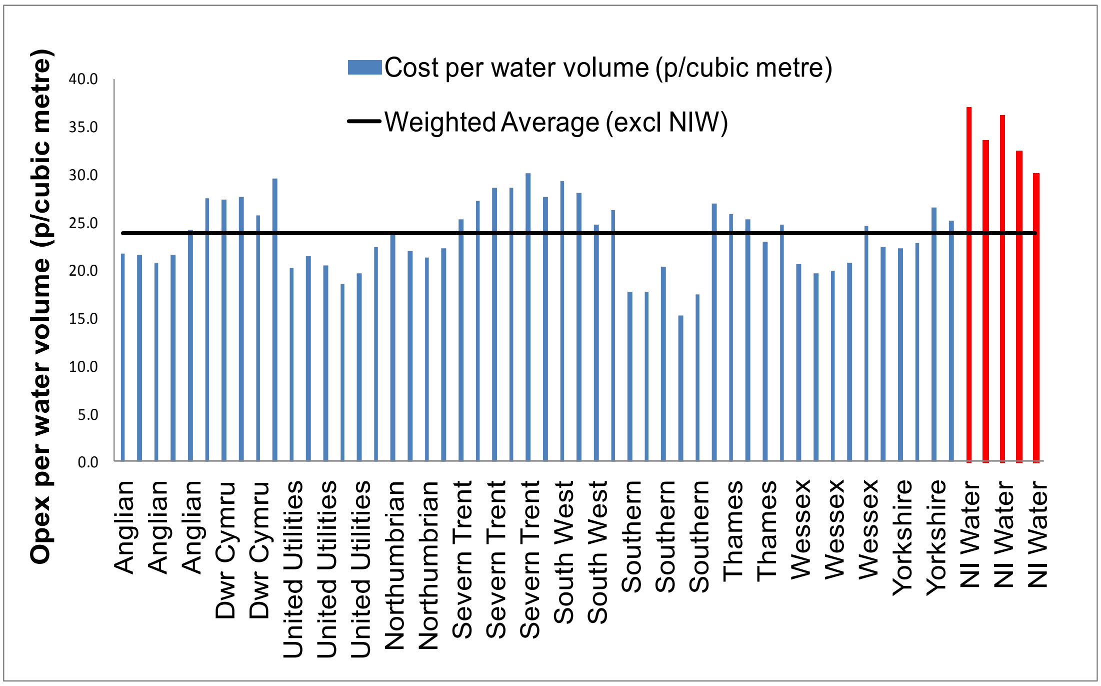
3.1.1 This approach mirrors the property method, the only difference being the denominator. In this instance, costs are divided by the volume of water entering the system. The form of the model is as follows:

Table 3.1 – Cost per water volume

Water Service:	Total opex (excluding business activities)
Data:	June Returns and PR14 August submission
Unit cost model:	The unit cost reflects the weighted industry opex cost per distribution input. Comparison is made of functional expenditure less service charges and business activities against predicted costs (distribution input multiplied by the weighted average industry unit cost).
pence /m³	Weighted average industry unit cost: 23.87p
Number of observations	50

Figures may not sum due to rounding

Figure 3.1 – Cost per water volume



3.1.2 The graph shows annual improvement for NI Water. These comparisons also show the company much closer to the industry average than before.

3.2. Sewage volumes

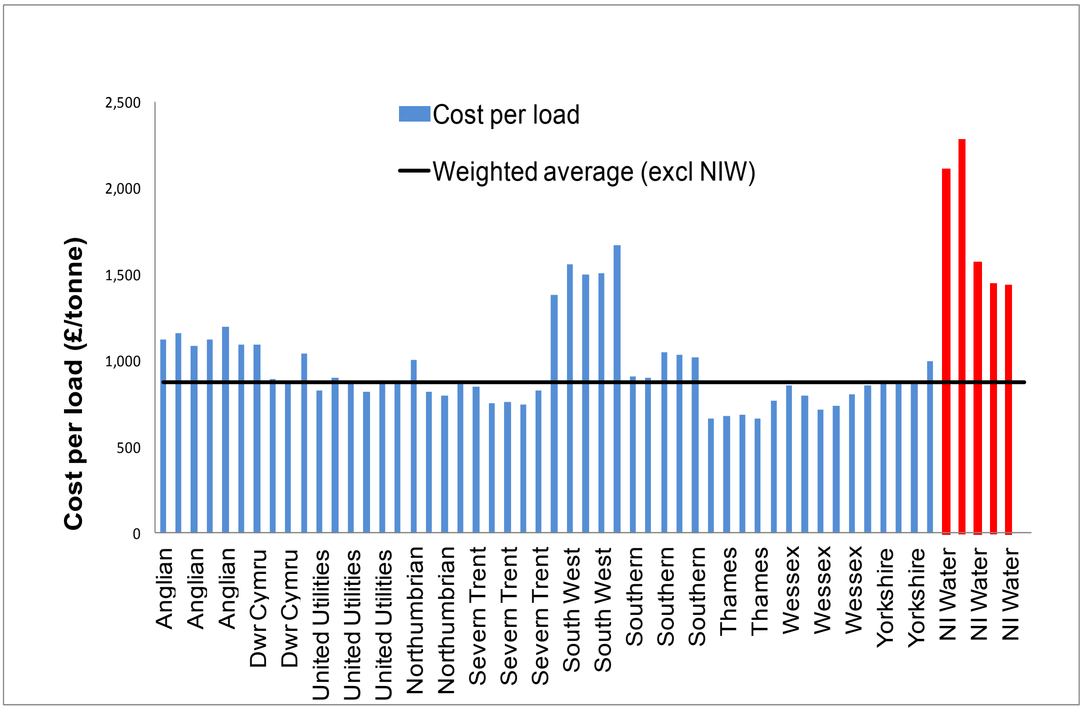
3.2.1 The total load entering the system explains sewage costs.

Table 3.2 – Cost per sewage load

Sewage Service:	Total opex (excluding business activities)
Data:	June Returns and PR14 August submission
Unit cost model:	The unit cost reflects the weighted industry opex cost per tonne of sewage load. Comparison is made of functional expenditure less service charges and business activities against predicted costs (the total load multiplied by the weighted average industry unit cost).
£/tonne BOD5	Weighted average industry unit cost: £872.13
Number of observations	50

Figures may not sum due to rounding

Figure 3.2 – Cost per sewage load



3.2.2 Whilst the sewage load has not varied greatly for NI Water over the years, costs have constantly been reducing.

3.3. Unit cost efficiency gap

3.3.1 In 2012-13 the efficiency gap using this unit cost is:

Table 3.3 – Volume unit cost efficiency gap for 2012-13

Efficiency Gap Calculation (2012-13 prices)					
	Category	**Process Rule**	**Water (£m)**	**Sewerage (£m)**	**Total (£m)**
A	NI Water actual cost		61.52	67.34	128.86
B	Less Adjustments		-3.01	-3.29	-6.30
C	**Modelled Cost**	**A – B**	**58.51**	**64.05**	**122.56**
D	Predicted Cost (average)		48.70	40.909	89.61
E	Difference	C – D	9.81	23.139	32.95
F	Adjustment Factor (%)		10%	20%	
G	Residual Adjustment	E * F	0.98	4.63	5.61
H	**New Predicted Costs**	**D + G**	**49.68**	**45.54**	**95.21**
I	Frontier Adjustment (%)		4.90%	-6.11%	
J	**Frontier Predicted Costs**	**H * (1 + I)**	**52.11**	**42.76**	**94.87**
K	Efficiency Gap (to average)	C – H	8.83	18.51	27.34
L	**Efficiency Gap % (to average)**	**K / C**	**15.10%**	**28.90%**	**22.31%**
M	Efficiency Gap (to frontier)	C – J	6.40	21.29	27.69
N	**Efficiency Gap % (to frontier)**	**M / C**	**10.94%**	**33.25%**	**22.60%**

Figures may not sum due to rounding.

3.3.2 The efficiency level is similar to the COLS findings. There does appear to be a slight anomaly on the water models, as the gap to the average is larger than the frontier. This illustrates the problem of maintaining the current benchmark companies.

3.3.3 Using volumes is arguably a better predictor than properties. Volumes measure actual activity so should be closely linked with cost. However, similar criticisms apply. No allowance is made for (dis)economies of scale and many of the explanatory variables are excluded.

4.0 Population Topex Regression

4.1. Water and sewage population

4.1.1 An alternative to the unit cost approach is a total opex (topex) regression. The benefit of regression is that scale is taken into consideration. Unlike unit costs, it can also account for the impact of more that one variable.

4.1.2 In this case, costs are modelled against the population served (water) and population connected (sewage). Any number of other variables could have been used e.g. properties, network size, volume etc. Results are below.

Table 4.1 – Water topex population model

Water Service:	**Total opex (excluding business activities)**	
Data:	June Returns and PR14 August submission	
Modelled cost:	ln (water functional expenditure less business activities, rates and service charges [£m])	
Explanatory Variables:	Coefficient	Standard Error
Constant	-4.265	0.346
Ln (population served)	1.054	0.042
Form of Model:	ln (modelled cost) = -4.265 + 1.054 * ln {population served}	
Statistical Indicators:	Nr. of observations = 50	R^2 = 0.930
	Standard error = 0.184	F test = 0.000

Table 4.2 – Sewage topex population model

Sewage Service:	**Total opex (excluding business activities)**	
Data:	June Returns and PR14 August submission	
Modelled cost:	ln (sewage functional expenditure less business activities, rates and service charges [£m])	
Explanatory Variables:	Coefficient	Standard Error
Constant	-1.664	0.309
Ln (connected population)	0.753	0.037
Form of Model:	ln (modelled cost) = -1.664 + 0.753 * ln {connected population}	
Statistical Indicators:	Nr. of observations = 50	R^2 = 0.898
	Standard error = 0.158	F test = 0.000

4.1.3 Both models give good statistical results. Explanatory variables are strongly significant. The regressions appear to be a good fit for the data as they suggest that population explains around 90% of opex spend.

4.2. Population topex model efficiency gap

4.2.1 Model results are given below:

Table 4.3 – Population model efficiency gap for 2012-13

Efficiency Gap Calculation (2012-13 prices)					
	Category	**Process Rule**	**Water (£m)**	**Sewerage (£m)**	**Total (£m)**
A	NI Water actual cost		61.52	67.34	128.86
B	Less Adjustments		-3.01	-3.29	-6.30
C	**Modelled Cost**	**A – B**	**58.51**	**64.05**	**122.56**
D	Predicted Cost (average)		38.42	46.80	85.22
E	Difference	C – D	20.09	17.25	37.34
F	Adjustment Factor (%)		10%	20%	
G	Residual Adjustment	E * F	2.01	3.45	5.46
H	**New Predicted Costs**	**D + G**	**40.43**	**50.25**	**90.68**
I	Frontier Adjustment (%)		0.61%	-19.62%	
J	**Frontier Predicted Costs**	**H * (1 + I)**	**40.67**	**40.39**	**81.06**
K	Efficiency Gap (to average)	C – H	18.08	13.80	31.88
L	**Efficiency Gap % (to average)**	**K / C**	**30.91%**	**21.55%**	**26.01%**
M	Efficiency Gap (to frontier)	C – J	17.84	23.66	41.50
N	**Efficiency Gap % (to frontier)**	**M / C**	**30.48%**	**36.94%**	**33.86%**

Figures may not sum due to rounding.

4.2.2 The result is a gap quite a bit larger than the COLS. It would be expected that population will influence cost. Population is however closely correlated with other key variables i.e. company size, usage, network connections etc. It is therefore not clear if population is the crucial factor.

4.2.3 Use of only one variable also fails to take proper account of particular company circumstances e.g. different network needs depending on how the population is distributed. This means that results must treated with caution.

5.0 CSV Topex Model A

5.1. Water CSV model A – equal weights

5.1.1 The composite scale variable (CSV) is an approach used by modellers. It is a very useful technique when there are a low number of data points. It is further helpful when there are a number of variables that might be expected to impact on costs, but are themselves highly correlated.

5.1.2 For the water service, the explanatory variable is a combination of mains length, population and distribution input. Each variable is given an equal weight (33/33/33) for ease of calculation. The model is in natural logs format. This helps predict the impact for both small and large companies.

5.1.3 Any number of variables or weights might have been employed. The chosen variables are already used in the COLS models and seemed the most obvious option.

5.1.4 Results are as follows:

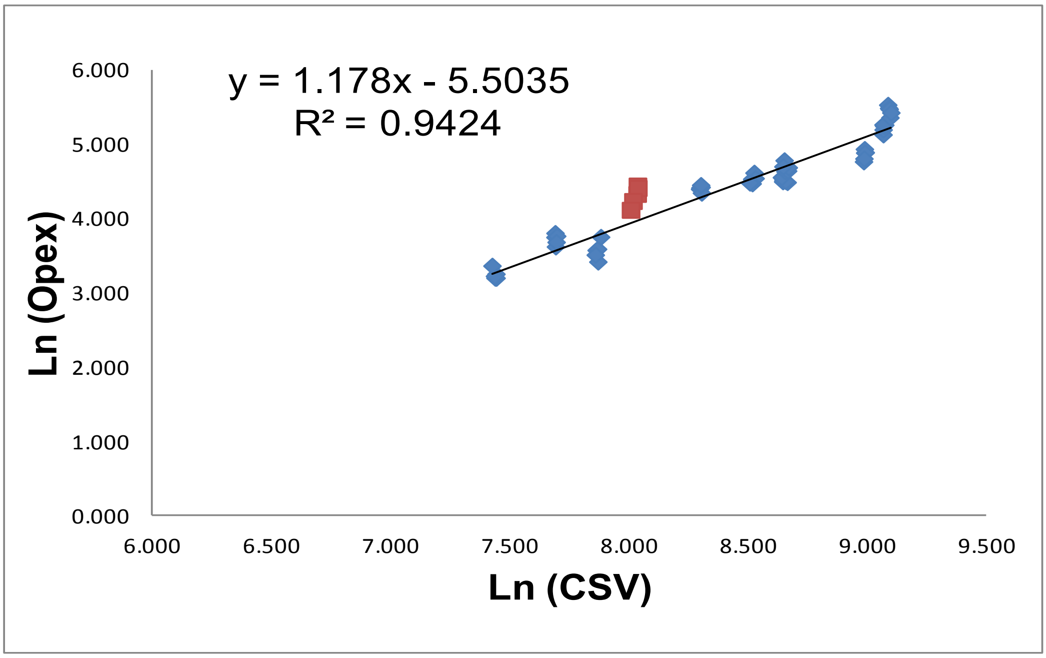
Table 5.1 – Water topex CSV model – equal weights

Water Service:	**Total opex (excluding business activities)**	
Data:	June Returns and PR14 August submission	
Modelled cost:	ln (water functional expenditure less business activities, rates and service charges [£m])	
Explanatory Variables:	Coefficient	Standard Error
Constant	-5.503	0.355
Ln (CSV)	1.178	0.042
Form of Model:	ln (modelled cost) = -5.503 + 1.178 * ln {CSV}	
Statistical Indicators:	Nr. of observations = 50	R^2 = 0.942
	Standard error = 0.166	F test = 0.000

5.1.5 This model is a very good fit for the data. The regression suggests that the composite variable explains almost 95% of water opex. As the model includes a number of variable impacts, findings are more robust than the single variable model.

5.1.6 In graphical form, the model looks as follows (NI Water observations in red):

Figure 5.1 – Water topex CSV model using equal weights



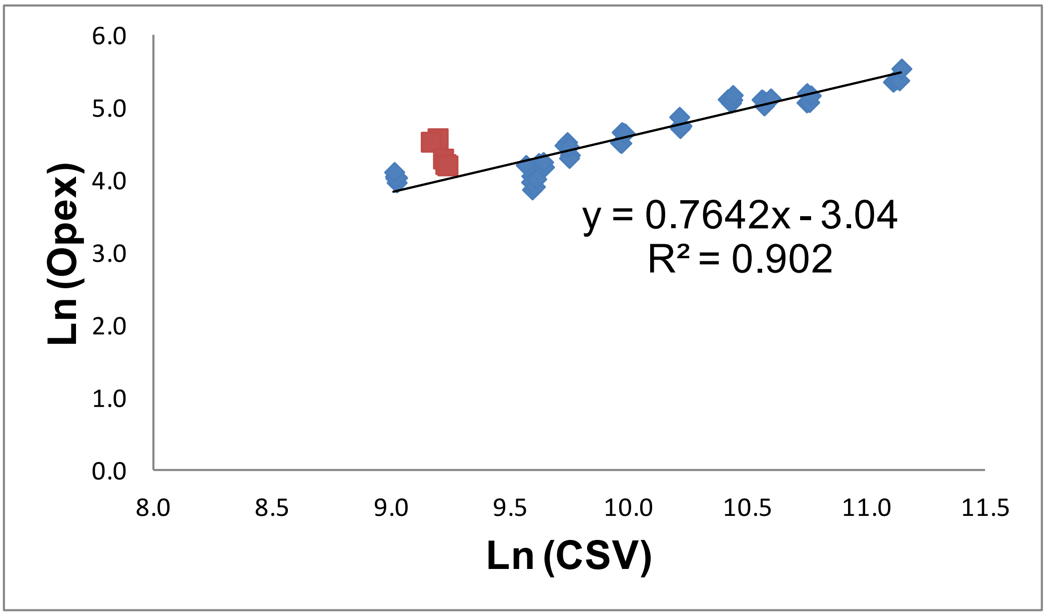
5.2. Sewage CSV model A – equal weights

5.2.1 The sewage composite variable is derived from the length of sewers, total load and the connected population. Again, all elements have an equal weighting.

Table 5.2 – Sewage topex CSV model – equal weights

Sewage Service:	**Total opex (excluding business activities)**	
Data:	June Returns and PR14 August submission	
Modelled cost:	ln (sewage functional expenditure less business activities, rates and service charges [£m])	
Explanatory Variables:	Coefficient	Standard Error
Constant	-3.040	0.368
Ln (CSV)	0.764	0.036
Form of Model:	ln (modelled cost) = -3.040 + 0.764 * ln {CSV}	
Statistical Indicators:	Nr. of observations = 50	R^2 = 0.902
	Standard error = 0.155	F test = 0.000

Figure 5.2 – Sewage topex CSV model – equal weights



5.2.2 The graphic illustrates NI Water in red. Whilst individual years cannot be observed, findings show the company moving toward average performance.

5.3. CSV topex model efficiency gap – equal weights

5.3.1 Model results are below:

Table 5.3 – CSV model A efficiency gap for 2012-13

Efficiency Gap Calculation (2012-13 prices)					
	Category	Process Rule	Water (%)	Sewerage (%)	Total (%)
A	Efficiency Gap % (to average)		11.74%	10.59%	11.14%
B	Efficiency Gap % (to frontier)		10.00%	28.93%	19.89%

Figures may not sum due to rounding.

5.3.2 The topex model gap is similar to the COLS findings, though slightly lower. Again, some adjustment may be required for water models as the imposed frontier is below average performance.

6.0 CSV Topex Model B

6.1. Water CSV model B – changed weights

6.1.1 The final model approach uses the same variables to generate a CSV. The only difference is that the weights are amended.

6.1.2 For water costs, there is a much higher correlation[2] with volumes and population than with network length. This may be due to the different sizes and location of mains, meaning they cannot be treated uniformly.

6.1.3 Since the correlation with cost is lower, the assumption is that this element of the CSV should be given a lesser weight. A 20/40/40 split has been adopted, with mains length given a reduced impact.

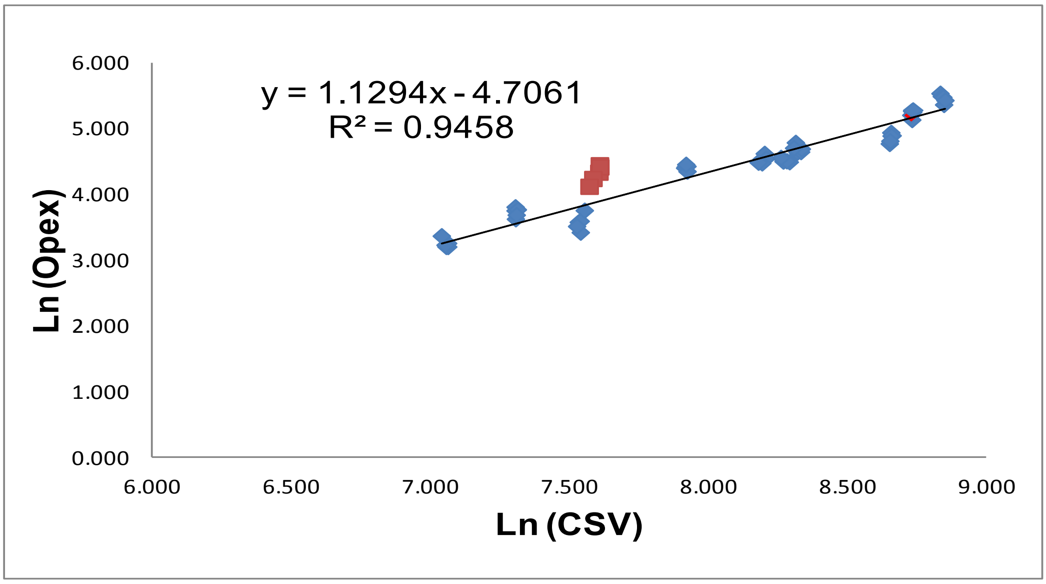
Table 6.1 – Water topex CSV model – different weights

Water Service:	**Total opex (excluding business activities)**	
Data:	June Returns and PR14 August submission	
Modelled cost:	ln (water functional expenditure less business activities, rates and service charges [£m])	
Explanatory Variables:	Coefficient	Standard Error
Constant	-4.706	0.316
Ln (CSV)	1.129	0.039
Form of Model:	ln (modelled cost) = -4.706 + 1.129 * ln {CSV}	
Statistical Indicators:	Nr. of observations = 50	R^2 = 0.946
	Standard error = 0.166	F test = 0.000

6.1.4 Whilst similar to the previous CSV model, statistical results are slightly improved. The goodness-of-fit of the variable suggests that omitted variables are not an issue. The improved results also indicate that mains length does have a lesser impact on costs.

[2] Reference here is to the Pearson correlation coefficient. This measures the strength of the linear relationship between two variables. Values range from -1 to 1 for negative and positive relationships respectively. A value close to zero indicates that a linear relationship does not exist.

Figure 6.1 – Water topex CSV model – different weights



6.2. Sewage CSV model B – changed weights

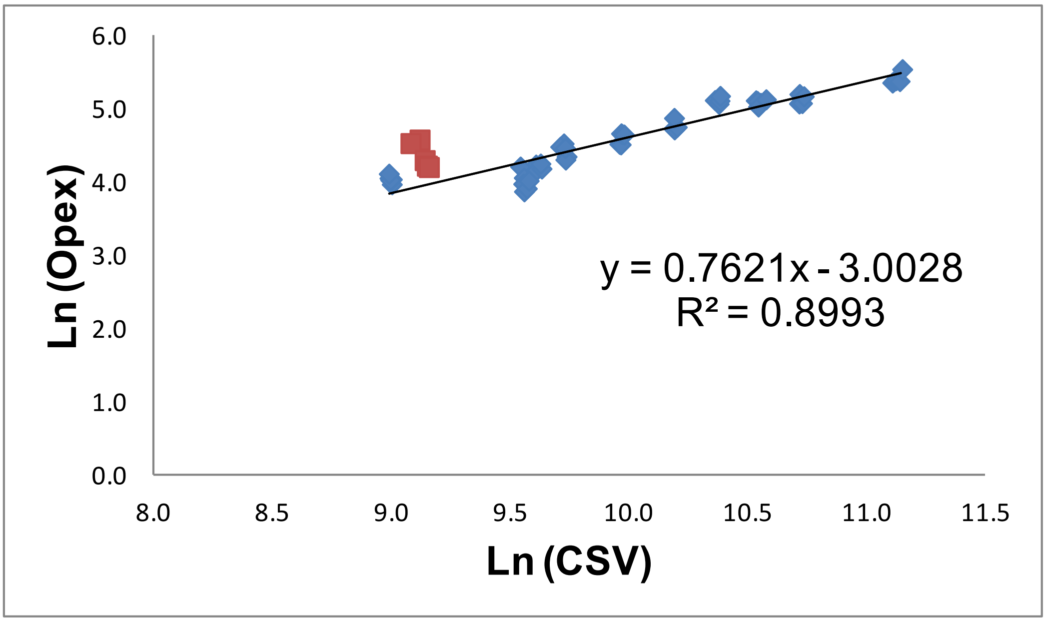
6.2.1 The difference in correlations does not exist on the sewerage models. The different split regression has been run for consistency purposes.

Table 6.2 – Sewage topex CSV model – different weights

Sewage Service:	**Total opex (excluding business activities)**	
Data:	June Returns and PR14 August submission	
Modelled cost:	ln (sewage functional expenditure less business activities, rates and service charges [£m])	
Explanatory Variables:	Coefficient	Standard Error
Constant	-3.003	0.372
Ln (CSV)	0.762	0.037
Form of Model:	ln (modelled cost) = -3.003 + 0.762 * ln {CSV}	
Statistical Indicators:	Nr. of observations = 50	R^2 = 0.899
	Standard error = 0.157	F test = 0.000

6.2.2 Results are very similar to the equal weight model. In this case, changing the weights has not improved estimations.

Figure 6.2 – Sewage topex CSV model – different weights



6.3. CSV topex model efficiency gap – different weights

6.3.1 Model results are below:

Table 6.3 – CSV model B efficiency gap for 2012-13

Efficiency Gap Calculation (2012-13 prices)					
	Category	Process Rule	Water (%)	Sewerage (%)	Total (%)
A	Efficiency Gap % (to average)		18.06%	13.40%	15.62%
B	Efficiency Gap % (to frontier)		15.96%	30.78%	23.71%

Figures may not sum due to rounding.

6.3.2 The topex model gap is very close to the COLS findings. Again, it can be seen that the frontier may be incorrect because of our imposed assumptions.

7.0 Summary of Findings

7.1. Comparing approaches

7.1.1 A summary of the model findings is given below.

Table 7.1 – Efficiency gap estimates using different methods

Methods	2012-13 Efficiency Gap (To Frontier)
COLS Models	23.2%
Unit Costs – (Properties)	36.3%
Unit Costs – (Volumes)	22.6%
Topex Regression (Population)	33.9%
Topex Regression A (CSV – 33/33/33)	19.9%
Topex Regression B (CSV - 20/40/40)	23.7%

7.1.2 Whilst more work is required to verify results, it would appear that the current COLS models falls within the range of efficiency results. In spite of a reasonably large degree of variability, it can be stated with certainty that a gap exists but has been falling over the past five years.

7.1.3 Calculation of the efficiency gap is not an exact science. This is why different models have been investigated.

7.1.4 Whilst useful in showing annual changes, the simple unit cost models are not considered the most robust. Even the population regression results are dubious given that not all costs can be explained adequately by a single variable.

7.1.5 The UR places most value on the CSV findings. What these show is that the efficiency gap estimate is close to the COLS results. They are perhaps even a little underestimated given the problems with the water frontier company.

7.1.6 Since findings are similar, the UR feel justified in using COLS to set efficiency targets. The CSV approach also uses updated cost and asset data. This helps provide assurance that the COLS approach of using old model comparisons has not introduced a bias that adversely affects NI Water.

7.1.7 Further work may be required. However, the alternative models help address company concerns and improve regulatory certainty by triangulating around a number of robust approaches.