

## Adverse impact on precision

### 1. Introduction

1. Section 5 of our provisional findings discusses whether the merger might be expected to reduce the precision of the econometric models from which Ofwat estimates technical efficiency targets, regardless of whether SEW or MKW is likely to form a benchmark. This appendix presents the results of some relevant analysis.
  
2. The outline of this appendix follows the structure of the discussion in the provisional findings. Section 2 discusses the precision of Ofwat's existing models: section 2(a) compares Ofwat's standard errors and bootstrapped standard errors and section 2(b) discusses Ofwat's ability to identify statistically significant relationships between costs and cost drivers. Section 3 discusses the impact of the merger on Ofwat's models: section 3(a) examines the effect of reducing the number of comparators in general on the basis of historical data and section 3(b) examines the effect of the loss of MKW and SEW as comparators and their replacement with a new, merged comparator again on the basis of historical data. There is an annex with some econometric results.

## 2. Precision of Ofwat’s existing models

### (a) Comparison between observed and bootstrapped standard errors

#### Operating expenditure

3. We compared the standard errors in Ofwat’s four econometric models for operating expenditure to alternative standard errors that we estimated with a statistical re-sampling technique called bootstrapping.<sup>1</sup> The results of our analysis are given in Table 1, which shows that Ofwat’s standard errors are biased but that the size of this bias is small compared with the standard errors (ie between 0.5 and 3.3 per cent). In fact, when the estimated bias is less than 25 per cent of the standard deviation of the bootstrapped estimate, the bias is not considered a problem.<sup>2</sup>

TABLE 1 Comparison of Ofwat’s standard errors for operating expenditure and bootstrapped standard errors

Model	Standard error			Bias	Bias as proportion (%) of ...	
	Ofwat	Bootstrapped*	CC Standard deviation†		Ofwat	Standard deviation
Water distribution	1.807	1.817	0.502	0.010	0.5	2.0
Resources & treatment:						
Cost driver 1	7.278	7.039	1.377	-0.239	3.3	17.4
Cost driver 2	2.734	2.703	0.748	-0.031	1.1	4.2
Power	0.028	0.028	0.006	0.000	1.2	5.4
Business activities	0.043	0.042	0.008	-0.001	3.0	16.0

Source: Ofwat and CC calculations.

\*Bootstrapped standard errors are based on 1,000 replications.

†The standard deviation of the bootstrapped estimate of the sampling distribution (known as the bootstrap distribution). This gives an estimate of the standard error of the unbiased standard error (sic).

4. Table 1 is interpreted as follows.

- The column labelled ‘Ofwat’ shows the standard errors (our measure of statistical precision) from Ofwat’s four econometric models of operating expenditure (ie water distribution, resources and treatment, power and business activities). Because the cost modelled in each model is measured in different units (because

<sup>1</sup>Using Ofwat’s 4 datasets of 22 observations for operating expenditure, we have randomly drawn with replacements 22 observations from each dataset. In each random draw, some of the original 22 observations appear once, some more than once and some not at all. We have then re-estimated Ofwat’s models and obtained the standard error of each coefficient (a measure of the width of the confidence interval). We have done this 1,000 times, each time drawing a new random sample and re-estimating the model. From this, we have 1,000 estimates of the standard errors for each model. From this data, we have estimated the standard error of the standard error (sic).

of each model's scaling factor<sup>3</sup>), and because each cost driver is different, the standard errors are not comparable across Ofwat's four models, which means the table is read row by row.

- The column labelled 'Bootstrapped' shows our estimate of Ofwat's standard errors for each model should be, were Ofwat to have more data (ie more than 22 comparators). These are generally close to Ofwat's standard errors of the estimates but not identical because Ofwat does not have an abundance of data.
- The column labelled 'Standard deviation' gives a measure of the precision of our estimate of what the standard error should be: these are very small (compared with the bootstrapped standard errors), which indicates that our estimates of what the standard error should be are very precise.<sup>4</sup>
- The column labelled 'Bias' is the difference between our bootstrapped standard errors (ie the 'Bootstrapped' column) and Ofwat's standard errors (ie the 'Ofwat' column).
- The column labelled 'Bias as a proportion of Ofwat' is the bias as a percentage of Ofwat's standard errors. Because the standard errors are not comparable from one Ofwat model to another (and so the bias is not comparable from one Ofwat model to another), this column gives a measure of the bias that can be compared across Ofwat models. This column tells us that Ofwat's standard errors are out by between 0.5 and 3.3 per cent.
- The final column 'Bias as a proportion of Standard deviation' is the bias as a percentage of the standard deviation of our bootstrapped estimate of the standard errors. Provided this is less than 25 per cent, the bias is not considered significant.

---

<sup>2</sup>See Efron, B (1982) *The Jackknife, the Bootstrap and other Resampling Plans*, Society for Industrial and Applied Mathematics: Philadelphia.

<sup>3</sup>The modelled cost for the water distribution model is expenditure on distribution per thousand head of resident winter population, the modelled cost for the resource and treatment model is expenditure on resource and treatment per million head of resident winter population, the modelled cost for the power model is simply expenditure on power, and the modelled cost for the business activities model is simply expenditure on business activities.

<sup>4</sup>This high degree of precision is because we used 1,000 replications in our bootstrapping procedure. Fewer replications would have resulted in less precision.

## Capital maintenance expenditure

5. We compared the standard errors in Ofwat's three econometric models for operating expenditure to alternative bootstrapped standard errors. The results of our analysis are given in Table 2, which shows that the standard errors of Ofwat's estimates are biased but that the size of this bias is small compared with Ofwat's standard errors (ie between 0.5 and 2.4 per cent) and compared with the standard deviation of the bootstrapped standard errors (ie between 1.7 and 7.7 per cent).

TABLE 2 Comparison of Ofwat's standard errors for capital maintenance expenditure and bootstrapped standard errors

Model	Standard error			Bias	Bias as proportion (%) of...	
	Ofwat	CC			Ofwat	Standard deviation
		Bootstrapped*	Standard deviation†			
Water distribution infrastructure	0.249	0.726	0.158	-0.008	1.1	4.9
Water distribution non-infrastructure	0.536	0.828	0.222	-0.004	0.5	1.7
Water management & general	0.234	10.335	3.247	-0.249	2.4	7.7

Source: Ofwat and CC calculations.

\*Bootstrapped standard errors are based on 1,000 replications.

†The standard deviation of the bootstrapped estimate of the sampling distribution (known as the bootstrap distribution). This gives an estimate of the standard error of the unbiased standard error (sic).

### **(b) Ofwat's ability to identify statistically significant relationships between costs and cost drivers**

6. In the course of constructing its econometric models, Ofwat used standard statistical tests to identify whether or not there were significant relationships between each of four operational expenditure cost measures, three capital maintenance cost measures and the corresponding cost drivers. Ofwat's ability to identify efficiency frontiers is contingent on its ability to identify statistically significant slopes in the OLS (ordinary least squares regression) models.

### *The power of statistical tests*

7. The probability of identifying a significant result is referred to as statistical power, which is connected to the notions of Type 1 and Type 2 errors. We explain these

concepts and examine the empirical evidence to assess Ofwat's ability to identify statistically significant results, both currently and if it were to lose a comparator.

8. Ofwat uses standard statistical tests to identify whether or not there are significant relationships between cost measures and cost drivers. It follows conventional practice and tests the overall significance of each model<sup>5</sup> and the individual significance of each of the components of the model.<sup>6</sup> The model of the relationship between cost and cost driver generally<sup>7</sup> consists of a line whose slope is determined empirically. Thus Ofwat's ability to explain the heterogeneity in water companies' costs relies on identifying a significant slope parameter for each model. Therefore we focus on tests to identify slope parameters that are significantly different from zero.

#### *Type 1 and Type 2 errors*

9. A statistical test can give a false result in one of two ways, known as errors of Type 1 and Type 2:
  - First, it may suggest that there is a significant slope when in fact no significant relationship exists: this is a Type 1 error.
  - Second, it may suggest that there is no significant slope when in fact a relationship does exist: this is a Type 2 error.
10. To guard against errors of Type 1, the criterion for significance is set sufficiently high so as to make false alarms infrequent. Ofwat knows that the relationship between cost drivers and cost should be positive (ie that the regression line slopes upward) so we set the probability of a Type 1 error equal to 2.5 per cent, so the test is of whether

---

<sup>5</sup>The overall significance is tested using the variance-ratio statistic which follows the F distribution under the assumptions of OLS regression.

<sup>6</sup>The significance of each parameter conditional on the presence of other parameters is tested using the ratio of the observed parameter to its standard error (SE). This follows the T distribution under the assumptions of OLS regression.

<sup>7</sup>Generally there is one cost driver. In the case of two cost drivers the relationship appears as a plane surface rather than a line.

the estimated relationship is positive when the real relationship is not.<sup>8</sup> That is, we assume the false alarm rate was one in 40.

11. However, a criterion set to achieve a low probability of Type 1 errors can increase the probability of Type 2 errors—the chance of missing a significant relationship. The ‘power function’ of the test measures the ability of the test to identify a significant relationship. In the case of OLS regression it is the probability of finding a significant slope as a function of the true slope. The probability of finding a significant result increases as the true slope gets larger (ie as the relationship between cost and driver strengthens).

#### *Empirical evidence on the power of Ofwat’s tests*

12. We examined the power of Ofwat’s tests empirically using the results of its OLS regressions as follows:
  - (a) First, we considered four values of slope parameter for each model:
    - (i) zero slope;
    - (ii) the observed OLS slope; and
    - (iii) the observed slope plus and minus its SE.
  - (b) Secondly, we computed the power of each test as if each of these four values were true.
  - (c) Thirdly, we repeated the calculations of power as if there were one less comparator.
13. Table 3 shows the results for operational expenditure models. It shows that Ofwat would certainly (ie with probability one) be expected to find a significant relationship if the true slope were the observed slope or the observed slope plus or minus one

---

<sup>8</sup>This corresponds to a one-tailed T-test at the 0.025 level. A two-tailed test at the 0.05 level uses the same criterion in the relevant tail.

standard error for both the power and business activities models, both with 22 and with 21 comparators. The resources and treatment model is reasonably robust to the loss of a comparator: provided that the true slope is at least equal to the observed slope the probability of identifying a significant result is over 72 per cent. In the distribution model, it is 52 per cent probable that if the true slope were equal to its observed value it would be detected as significant. If a comparator were lost this probability declines to just below 50 per cent.

TABLE 3 Statistical power of Ofwat's statistical tests for 2005/06 operating expenditure models

<i>Model</i>	<i>Cost driver</i>	<i>Number of comparators Count</i>	<i>Magnitude of slope under alternative hypothesis</i>			
			<i>Zero Slope* Power†</i>	<i>Observed Slope - SE Power†</i>	<i>Observed Slope Power†</i>	<i>Observed Slope + SE Power†</i>
Power	Log (Distribution Input x Average Pumping Head)	22	0.025	1.000	1.000	1.000
		21	0.025	1.000	1.000	1.000
Distribution	Proportion of large mains	22	0.025	0.178	0.522	0.848
		21	0.025	0.163	0.497	0.834
Business activities	Log (number of billed properties)	22	0.025	1.000	1.000	1.000
		21	0.025	1.000	1.000	1.000
Resources & treatment	Total number of sources/DI	22	0.025	0.642	0.906	0.986
		21	0.025	0.604	0.890	0.982
	Proportion of DI from Boreholes	22	0.025	0.378	0.749	0.946
		21	0.025	0.348	0.723	0.937

Source: Ofwat, CC calculations.

\*Assumes Ofwat tests either for a significant positive slope or a significant negative slope at the 2.5 per cent level of significance: a 'one-tailed t-test'.

†The statistical power is the probability of detecting a significant slope of the OLS regression function, if the true slope were the given slope. Thus, in the case where the true slope is actually zero, there is a 2.5 per cent probability of detecting a significant slope erroneously; this is the probability of a Type 1 error and equals the level of significance of the test.

14. Table 4 shows robust results for capital maintenance expenditure models. The water distribution non-infrastructure model is robust to the loss of a comparator because Ofwat should have at least a 90 per cent probability of detecting a significant relationship even if it loses one comparator. This is evident since the statistical power is at least 0.90, as long as the true slope is at least equal to the observed slope minus one standard error. The water distribution infrastructure and management and general models are reasonably robust. Provided that the true slope is equal to the observed slope or more, the probability of identifying significant results is over 71 per cent.

TABLE 4 Statistical power of Ofwat's statistical tests for 2005/06 capital maintenance expenditure models

Model	Cost driver	Number of comparators Count	Magnitude of slope under alternative hypothesis			
			Zero Slope* Power†	Observed Slope – SE Power†	Observed Slope Power†	Observed Slope + SE Power†
Water distribution non-infrastructure	Log (storage capacity/pumping station capacity)	22	0.025	0.922	0.989	0.999
		21	0.025	0.904	0.985	0.998
Water distribution infrastructure	Log (connected properties per length of main)	22	0.025	0.476	0.820	0.967
		21	0.025	0.443	0.798	0.960
Water management & general	Proportion of properties that are non-household	21‡	0.025	0.371	0.743	0.944
		20‡	0.025	0.341	0.716	0.935

Source: Ofwat, CC calculations.

\*Assumes Ofwat tests either for a significant positive slope or a significant negative slope at the 2.5 per cent level of significance: a 'one-tailed t-test'.

†The statistical power is the probability of detecting a significant slope of the OLS regression function, if the true slope were the given slope. Thus, in the case where the true slope is actually zero, there is a 2.5 per cent probability of detecting a significant slope erroneously; this is the probability of a Type 1 error and equals the level of significance of the test.

‡Excludes Portsmouth Water.

### 3. Impact of the merger

#### ***(b) The effect of reducing the number of comparators generally, on the basis of historical data***

##### *Operating expenditure*

15. For each of Ofwat's four econometric models of operating expenditure, we checked the sensitivity of the results to errors in the slope of the efficiency frontier. We proceeded in a number of steps:

- First, we found the total cost difference of all companies from the efficiency frontier for each model.
- Secondly, we found the total cost difference of all companies from a frontier whose slope was reduced by the error in the slope.
- Thirdly, we found the total cost difference of all companies from a frontier whose slope was increased by the error in the slope.
- Fourthly, we compared the total cost from the second and third steps with the total cost from the first step and computed the average deviation in cost.
- Next, we computed the error in the slope that would be expected if there were one less comparator. With this revised error we repeated the four steps to arrive at a new average deviation in cost.
- Finally, we compared the average deviations to find the marginal rate of increase in confidence interval due to the loss of a comparator.

16. Our results are shown in Table 5.

TABLE 5 Summary of sensitivity analysis on confidence intervals for 2005–06 operating expenditure models

<i>Model</i>	<i>Model sensitivity to slope</i>			<i>Mean deviation</i> £m	<i>Change in deviation</i>	
	<i>Slope – SE</i> £m	<i>Current slope</i> £m	<i>Slope + SE</i> £m		£m	%
<b>2005–06 Power model</b>						
<i>With current residual degrees of freedom</i>						
Total cost difference of all companies from frontier	89.1	108.4	128.6	19.7		
Benchmark	SVT	SVT	SVT			
<i>With reduced residual degrees of freedom</i>						
Total cost difference of all companies from frontier	88.6	108.4	129.1	20.2	0.5	2.6
Benchmark	SVT	SVT	SVT			
<b>2005–06 Resources and treatment model*</b>						
<i>With current residual degrees of freedom</i>						
Total cost difference of all companies from frontier	348.5	81.7	141.2	163.2		
Benchmark	UUW	SRN	SRN			
<i>With reduced residual degrees of freedom</i>						
Total cost difference of all companies from frontier	352.2	81.7	142.8	165.8	2.7	1.6
Benchmark	UUW	SRN	SRN			
<b>2005–06 Water distribution</b>						
<i>With current residual degrees of freedom</i>						
Total cost difference of all companies from frontier	308.9	394.1	484.8	88.0		
Benchmark	NWT	NWT	NWT			
<i>With reduced residual degrees of freedom</i>						
Total cost difference of all companies from frontier	306.8	394.1	487.3	90.2	2.3	2.6
Benchmark	NWT	NWT	NWT			
<b>2005–06 Business activities model</b>						
<i>With current residual degrees of freedom</i>						
Total cost difference of all companies from frontier	618.8	205.8	738.4	472.8		
Benchmark	YKY	YKY	YKY			
<i>With reduced residual degrees of freedom</i>						
Total cost difference of all companies from frontier	617.4	205.8	740.0	472.9	0.1	0.0
Benchmark	YKY	YKY	YKY			

Source: Ofwat, CC calculations.

\*This model has two slope parameters. In the 'Slope – SE' column, both are reduced by SE. In the 'Slope + SE' column, both are increased by SE.

### *Capital maintenance expenditure*

17. We checked the sensitivity of the results to errors in the slope of the efficiency benchmark for each of Ofwat's three econometric models of capital maintenance expenditure, using the methodology described above. Our results are shown in Table 6.

TABLE 6 Summary of sensitivity analysis on confidence intervals for 2005/06 capital maintenance expenditure models

<i>Model</i>	<i>Model sensitivity to slope</i>			<i>Mean deviation £m</i>	<i>Change in deviation</i>	
	<i>Slope – SE £m</i>	<i>Current slope £m</i>	<i>Slope + SE £m</i>		<i>£m</i>	<i>%</i>
<i>2005–06 Water management &amp; general model</i>						
<i>With current residual degrees of freedom</i>						
Total cost difference of all companies from frontier	95.9	85.5	75.3	10.3		
Benchmark	NES	NES	NES			
<i>With reduced residual degrees of freedom</i>						
Total cost difference of all companies from frontier	96.1	85.5	75.0	10.6	0.3	2.7
Benchmark	NES	NES	NES			
<i>2005–06 Water distribution infrastructure model</i>						
<i>With current residual degrees of freedom</i>						
Total cost difference of all companies from frontier	142.2	107.4	115.4	21.4		
Benchmark	ANH	NES	NES			
<i>With reduced residual degrees of freedom</i>						
Total cost difference of all companies from frontier	143.2	107.4	115.6	22.0	0.6	2.9
Benchmark	ANH	NES	NES			
<i>2005–06 Water distribution non-infrastructure model</i>						
<i>With current residual degrees of freedom</i>						
Total cost difference of all companies from frontier	111.5	79.4	54.3	28.6		
Benchmark	tvn	tvn	tvn			
<i>With reduced residual degrees of freedom</i>						
Total cost difference of all companies from frontier	112.4	79.4	53.7	29.4	0.8	2.7
Benchmark	tvn	tvn	tvn			

Source: Ofwat, CC calculations.

**(b) The effect of the loss of MKW and SEW and their replacement with a merged comparator**

*Operating expenditure*

18. We compared the output of four operating expenditure models Ofwat's models with 22 observations against those obtained from the alternative model using 21 observations, ie with a merged MKW/SEW. The detailed results of our analysis are given in the annex.
19. Table 7 expresses some results of the analysis described in the paragraph above in the same format as Table 5. That is, it compares the average cost difference per company from the frontier with and without the merger.

TABLE 7 **Effect of the merger on the average cost difference per company from the frontier for 2005/06 operating expenditure models**

<i>Model</i>	<i>Current*</i> £m	<i>With merger†</i> £m	<i>Deviation</i>	
			£m	%
<b>2005/06 Power model</b>				
Total cost difference of all companies from frontier	108.4	106.8		
Cost difference from frontier per company	4.9	5.1	0.2	3.2
Equivalent cost difference of all companies from frontier‡	108.4	111.9	3.5	
Benchmark	SVT	SVT		
<b>2005/06 Resources and treatment model</b>				
Total cost difference of all companies from frontier	81.7	76.5		
Cost difference from frontier per company	3.7	3.6	-0.1	-1.9
Equivalent cost difference of all companies from frontier‡	81.7	80.1	-1.6	
Benchmark	SRN	SRN		
<b>2005/06 Water distribution</b>				
Total cost difference of all companies from frontier	394.1	382.3		
Cost difference from frontier per company	17.9	18.2	0.3	1.6
Equivalent cost difference of all companies from frontier‡	394.1	400.5	6.3	
Benchmark	NWT	NWT		
<b>2005/06 Business activities model</b>				
Total cost difference of all companies from frontier	205.8	197.6		
Cost difference from frontier per company	9.4	9.4	0.1	0.6
Equivalent cost difference of all companies from frontier‡	205.8	207.0	1.2	
Benchmark	YKY	YKY		

Source: Ofwat, CC calculations.

\*With 22 comparators.

†With 21 comparators.

‡Scaling up loss per company with 21 comparators as if there were 22 comparators.

## Capital maintenance

20. We compared the output of three capital maintenance expenditure models Ofwat's models with 22 observations against those obtained from the alternative model using 21 observations, ie with a merged MKW/SEW. The detailed results of our analysis are given in the annex.
21. Table 8 expresses some results of the analysis described in the paragraph above in the same format as Table 6. That is, it compares the average cost difference per company from the frontier with and without the merger.

TABLE 8 **Effect of the merger on average cost difference per company from the frontier for 2005/06 capital maintenance expenditure models**

<i>Model</i>	<i>Current*</i>	<i>With</i>	<i>Deviation</i>	
	<i>£m</i>	<i>merger†</i>	<i>£m</i>	<i>%</i>
<b>2005/06 Water Distribution Infrastructure</b>				
Total cost difference of all companies from frontier	107.4	101.8		
Cost difference from frontier per company	4.9	4.8	0.0	-0.7
Equivalent cost difference of all companies from frontier‡	107.4	106.6	-0.8	
Benchmark	NES	NES		
<b>2005/06 Water Distribution Non-Infrastructure</b>				
Total cost difference of all companies from frontier	79.4	74.8		
Cost difference from frontier per company	3.6	3.6	0.0	-1.2
Equivalent cost difference of all companies from frontier‡	79.4	78.4	-1.0	
Benchmark	tvn	ses		
<b>2005/06 Water Management and General</b>				
Total cost difference of all companies from frontier	85.5	78.3		
Cost difference from frontier per company	3.9	3.7	-0.2	-4.1
Equivalent cost difference of all companies from frontier‡	85.5	82.0	-3.5	
Benchmark	NES	NES		

Source: Ofwat, CC calculations.

\*With 22 comparators.

†With 21 comparators.

‡Scaling up loss per company with 21 comparators as if there were 22 comparators.

### **Analysis of the impact of the merger on Ofwat's models using historical data**

1. We analyse the impact of the merger between SEW and MKW on the robustness and precision of the econometric models that Ofwat currently uses to model relative efficiency. To put the results of this analysis into context, we assess the extent to which Ofwat's models change over time.
  
2. The remainder of this annex is structured as follows. Section 1 discusses the measures we use to assess robustness (section 1(a) discusses qualitative robustness and section 1(b) discusses quantitative robustness). Section 2 discusses variation over time in the results of Ofwat's models for operating expenditure (section 2(a)) and capital maintenance expenditure (section 2(b)). Section 3 reports the results of our simulation of the effects of the merger of MKW and SEW on Ofwat's operating expenditure models (section 3(a)) and on Ofwat's capital maintenance expenditure models (section 3(b)). Finally section 4 compares the results of sections 2 and 3: that is, it compares changes over time with changes from the merger in the results of Ofwat's models for operating expenditure (section 4(a)) and capital maintenance expenditure (section 4(b)).

#### **1. Robustness**

3. Robustness properties fall within two frameworks: qualitative and quantitative robustness. Qualitative robustness focuses on stability and performance loss over a family of slightly deviating distributions of the underlying data. Quantitative robustness considers the sensitivity of estimators to a proportion of aberrant observations.

### **(a) Qualitative robustness**

4. Ofwat considers the following statistics, all of which are suitable for assessing qualitative robustness:

(a) Model significance (F test): Ofwat uses this test statistic to determine whether there is an association between the modelled cost and cost driver(s). The numbers in the tables we present in the current analysis are the probability that there is no association between the variables, so that a value less than 5 per cent indicates a statistically significant association between cost and cost drivers.

(b)  $R^2$ : This statistic shows how closely the variables in the model are related to each other. It lies between 0 and 1. In the current context, an  $R^2$  of 1 means that all the variation in the cost variable is explained by the variation in the cost driver.<sup>1</sup> This statistic is based on the variation in costs and cost drivers. Removing one comparator reduces the variation in both variables. For this reason,  $R^2$  of models that include a different number of observations are not strictly comparable. We therefore use a closely related, alternative test statistic called 'adjusted  $R^2$ ' ( $\bar{R}^2$ ). The intuitive interpretation of both statistics is the same.<sup>2</sup>

(c) Standard errors of models: This is a measure of the overall variation in a model. It measures how spread out the data is around the regression line. The further away the data points are from the regression line the greater the value of the standard error.

---

<sup>1</sup>An  $R^2$  value of 1 implies that expenditures are perfectly predictable by the econometric models and subsequently that all companies are perfectly efficient. However, Ofwat's econometric models are not designed to predict efficiency; rather, by having diversity on performance, the models allow Ofwat to identify which business units are relatively inefficient. In the current context, a relatively low value of  $R^2$  therefore does not imply that a model performs poorly. An alternative interpretation would be that the diversity in performance is large.

<sup>2</sup>The loss of a data point (ie a comparator) results in a reduction in the degrees of freedom. The measure of adjusted  $R^2$  ( $\bar{R}^2 = 1 - (1 - R^2)[(n - 1)/(n - k - 1)]$ ) takes this change into account and should therefore be the preferred measure for comparison purposes.

5. In addition to these statistics we consider estimates of cost driver's coefficients and the confidence intervals around them. Estimated coefficients quantify the effect of a cost driver on costs. A confidence interval indicates the precision of these estimates.

**(b) Quantitative robustness**

6. To assess the impact of the loss of a comparator on the results of an econometric model, we analyse Cook's distances pre- and post-merger. This statistic measures the influence of a data point on the results of an econometric model. Observations with high leverage influence disproportionately the results of an econometric model.<sup>3</sup>

**2. Variation in the results of Ofwat's econometric models over time**

7. To put the results of our analysis of the effects of the merger into context, in this section we assess the extent to which Ofwat's models change over time absent the merger. To this end we consider the following measures: coefficient estimates, width of confidence intervals (measured in terms of proportion of the estimate), and adjusted R squared.
8. The results of Ofwat's econometric models may change over times for three main reasons:
  - because costs and cost drivers change over time;
  - because Ofwat changes the way it calculates companies' costs and cost drivers;  
and
  - because Ofwat changes the cost driver.
9. Indeed, Ofwat changed the way it calculates the cost variables in its three capital maintenance expenditure models every year between 2002/03 and 2005/06. All

---

<sup>3</sup>Cook's distances are relative measures that can be compared across models.

capital maintenance cost variables are averages calculated over a number of years. In 2002/03, Ofwat calculated average capital expenditure over a period of five years. This number changed to six years in 2003/04 and seven years in 2004/05. In 2005/06, Ofwat calculated average expenditure in its capital maintenance models over a period of six years.

10. Further, Ofwat changed the cost driver in one of its operating expenditure models. The two cost drivers in the resource and treatment cost model previously were (a) the number of sources divided by distribution input (ML/d), and (b) the proportion of supplies derived from river sources in previous years. In its 2005/06 report, Ofwat changed the latter cost driver to ‘proportion of supplies derived from boreholes’.<sup>4</sup>

### **(a) Operating expenditure**

11. Table 1 lists estimates of cost drivers’ coefficients from Ofwat’s four econometric operating expenditure models between 2002/03 and 2005/06.

TABLE 1 **Variation over time in cost drivers’ point estimates from Ofwat’s econometric models of operating expenditure**

Cost	Point estimates of cost drivers’ coefficients			
	2002/03	2003/04	2004/05	2005/06
Distribution	13.27	6.11	5.01	4.53
Power	0.94	0.93	0.93	0.91
Business activities	0.95	0.92	0.93	0.95
Resource and treatment (driver A)	16.77	22.42	18.48	23.85
Resource and treatment (driver B)	5.12	5.93	4.13	N/A

Source: Ofwat and CC calculations.

12. Table 1 illustrates the extent to which the estimated coefficients change over time. For example, in the distribution model, the cost driver’s coefficient varies between 13.3 (2002/03) and 4.5 (in 2005/06). Similarly, the point estimate of cost driver A in

<sup>4</sup>These two variables are negatively correlated, that is the higher the proportion of supplies derived from river sources, the smaller the proportion of supplies derived from boreholes.

the resource and treatment model varies between 16.8 (in 2002/03) and 23.9 (in 2005/06).

13. The point estimate of cost driver B in the resource and treatment model is not available for the year 2005/06. This is because Ofwat replaced this cost driver with an alternative one (see paragraph 10).
14. Table 1 also illustrates that point estimates of cost drivers' coefficients in the models operating expenditure econometric models 'power' and 'business activities' are more stable over time.
15. Table 2 lists the confidence intervals around the estimates of cost drivers' coefficients for Ofwat's econometric operating expenditure models between 2002/03 and 2005/06. The confidence intervals are expressed in terms of percentages of the corresponding point estimate.

TABLE 2 **Variation over time in confidence intervals around cost drivers' point estimates from Ofwat's econometric models of operating expenditure**

Cost	<i>Confidence intervals as percentage of the estimate</i>			
	2002/03	2003/04	2004/05	2005/06
Distribution	123.5	66.3	75.2	97.4
Power	5.0	5.5	6.2	5.4
Business activities	8.7	9.5	9.7	7.3
Resource and treatment (driver A)	78.2	61.2	82.4	59.9
Resource and treatment (driver B)	100.0	87.7	138.6	74.2

Source: Ofwat and CC calculations.

16. Table 2 illustrates that the width of the confidence intervals vary considerably over time. For example, the width of the confidence interval in the distribution model ranges from 66 per cent in 2003/04 to 124 per cent in the previous year.<sup>5</sup>
17. In contrast, the width of the confidence intervals in the 'power' and 'business activities' models are tighter with 5 to 6 per cent and 7 to 10 per cent, respectively. The width of the confidence intervals also varies to lesser extent, compared with those in the 'distribution' model.
18. Table 3 lists values of 'adjusted R squared', a measure of goodness of fit, for Ofwat's econometric operating expenditure models between 2002/03 and 2005/06.

TABLE 3 Variation over time in the 'goodness of fit' measure (adjusted R squared) from Ofwat's econometric models of operating expenditure

<i>Cost</i>	<i>Adjusted R squared</i>			
	<i>2002/03</i>	<i>2003/04</i>	<i>2004/05</i>	<i>2005/06</i>
Distribution	0.36	0.30	0.24	0.15
Power	0.99	0.99	0.98	0.99
Business activities	0.96	0.96	0.96	0.98
Resource and treatment (driver A)	0.20	0.32	0.18	0.33
Resource and treatment (driver B)	0.36	0.30	0.24	0.15

Source: Ofwat and CC calculations.

19. Table 3 illustrates that this measure of goodness of fit varies considerably in most models. For example, the adjusted R squared value for the operating expenditure model 'distribution' varies between 0.15 (in 2005/06) and 0.36 (in 2002/03). In contrast, the adjusted R squared values for the operating expenditure models 'power' and 'business activities' stay relatively constant between 0.98 (in 2004/05) and 0.99 in all other years, and 0.98 in 0.96 in all other years, respectively.

<sup>5</sup>A confidence interval's width in excess of 100 per cent indicates that the estimated coefficient is not statistically significantly different from zero at the chosen level of significance/certainty (here 95 per cent).

**(b) Capital maintenance expenditure**

20. Table 4 lists estimated coefficients of cost drivers from Ofwat's three econometric capital maintenance expenditure models between 2002/03 and 2005/06.

**TABLE 4 Variation over time in cost drivers' point estimates from Ofwat's econometric models of capital maintenance expenditure**

<i>Cost</i>	<i>Point estimates of cost drivers' coefficients</i>			
	<i>2002/03</i>	<i>2003/04</i>	<i>2004/05</i>	<i>2005/06</i>
Distribution infrastructure	0.89	0.76	0.73	0.74
Distribution non-infrastructure	0.66	0.74	0.83	0.94
Management and general	13.02	12.77	10.33	9.17

*Source:* Ofwat and CC calculations.

---

21. Table 4 illustrates the degree by which point estimates of cost drivers coefficients change in Ofwat's models over time.

22. The cost drivers' estimated coefficients for capital maintenance models vary considerably over time. For example, in case of the model 'management and general', point estimates vary between 9.2 (in 2005/06) and 13.0 (in 2002/03).

23. Table 5 lists the confidence intervals around cost drivers' coefficients' point estimates for Ofwat's capital maintenance econometric models between 2002/03 and 2005/06. The confidence intervals are expressed in terms of percentages of the corresponding point estimate.

**TABLE 5 Variation over time in confidence intervals around cost drivers' point estimates from Ofwat's econometric models of capital maintenance expenditure**

<i>Cost</i>	<i>Confidence intervals as percentage of the estimate</i>			
	<i>2002/03</i>	<i>2003/04</i>	<i>2004/05</i>	<i>2005/06</i>
Distribution infrastructure	47.1	61.6	67.2	69.0
Distribution non-infrastructure	65.2	56.3	49.2	45.7
Management and general	93.2	90.1	75.5	75.9

*Source:* Ofwat and CC calculations.

---

24. Table 5 illustrates the degree to which the width of confidence intervals of Ofwat's econometric capital maintenance models varies over time. The confidence intervals of the three capital maintenance models vary in similar proportions. For example, the confidence interval of the distribution infrastructure model varies between 47 and 69 per cent between 2002/03 and 2005/06.

25. Table 6 lists values of 'adjusted R squared', a measure of goodness of fit,<sup>6</sup> for Ofwat's capital maintenance expenditure models between 2002/03 and 2005/06.

TABLE 6 Variation over time in the 'goodness of fit' measure (adjusted R squared) from Ofwat's econometric models of capital maintenance expenditure

Cost	Adjusted R squared			
	2002/03	2003/04	2004/05	2005/06
Distribution infrastructure	0.47	0.33	0.29	0.28
Distribution non-infrastructure	0.31	0.38	0.45	0.49
Management and general	0.16	0.17	0.25	0.25

Source: Ofwat and CC calculations.

26. Table 6 illustrates that this measure of goodness of fit varies considerably in the case of most models.

### 3. Simulating a merger between MKW and SEW

#### (a) Operating expenditure

27. Table 7 compares the results of Ofwat's operating expenditure econometric modelling against the results after merging MKW and SEW to one company (ie moving from 22 to 21 observations).<sup>7</sup> The table allows comparisons between (a) probability values of F-tests, (b) model standard errors, and (c) adjusted  $R^2$ .

<sup>6</sup>Adjusted R squared values lie between 0 and 1 and measure the proportion of variation in the cost that is explained by variation in the cost driver.

<sup>7</sup>We created the data point for the merged entity MKW and SEW by adding or averaging the data for the two companies, as appropriate.

TABLE 7 The results of Ofwat’s econometric modelling of operating expenditure pre- and post-merger

	<i>F-tests (p-value (%))</i>		<i>Model standard errors</i>		<i>Adjusted R<sup>2</sup></i>	
	<i>Pre-merger</i>	<i>Post-merger</i>	<i>Pre-merger</i>	<i>Post-merger</i>	<i>Pre-merger</i>	<i>Post-merger</i>
Distribution model	1.2	2.1	0.2	0.2	0.24	0.21
Power model	0.0	0.0	0.2	0.2	0.98	0.98
Business activities model	0.0	0.0	0.2	0.3	0.96	0.96
Resource and treatment	6.0	5.7	2.3	2.3	0.18	0.19

Source: Ofwat, CC analysis.

28. Table 8 compares estimates of cost drivers’ parameters and their confidence intervals.

TABLE 8 Coefficients from Ofwat’s econometric models of operating expenditure pre- and post-merger

	<i>Point estimates of cost drivers’ coefficients</i>		<i>Confidence intervals around estimates (%)</i>	
	<i>Pre-merger</i>	<i>Post-merger</i>	<i>Pre-merger</i>	<i>Post-merger</i>
Distribution model	5.0	5.0	75.2	82.9
Power model	0.9	0.9	6.2	6.2
Business activities model	0.9	0.9	9.7	9.9
Resource and treatment	18.5	19.4	82.4	81.5
Resource and treatment (second cost driver)	4.1	4.0	138.6	145.8

Source: Ofwat, CC analysis.

29. Table 9 compares Cook’s distances from Ofwat’s operating expenditure econometric models pre- and post-merger.

TABLE 9 Cook’s distances from Ofwat’s econometric models of operating expenditure pre- and post-merger

	<i>Average</i>		<i>Maximum</i>	
	<i>Pre-merger</i>	<i>Post-merger</i>	<i>Pre-merger</i>	<i>Post-merger</i>
Distribution model	0.04	0.05	0.42	0.41
Power model	0.04	0.05	0.15	0.14
Business activities model	0.06	0.06	0.51	0.55
Resource and treatment	0.07	0.09	0.79	0.85

Source: Ofwat, CC analysis.

30. Table 9 shows that, on average, average Cook’s distances increase as the number of comparators decreases. This means that post-merger, every observation has a

greater effect on the results of the regression analysis than pre-merger. This follows naturally from the reduced number of observations post-merger. Table 9 does not indicate that there is an individual leverage point in any of the models, pre- or post-merger.

**(b) Capital maintenance expenditure**

31. Table 10 compares the results of Ofwat’s capital maintenance econometric modelling against the results after merging MKW and SEW to one company (ie moving from 22 to 21 observations). The table allows comparisons between (a) probability values of F-tests, (b) model standard errors, and (c) adjusted  $R^2$ .

TABLE 10 The results of Ofwat’s econometric models of capital maintenance expenditure pre- and post-merger

	<i>F-tests (p-value (%))</i>		<i>Model standard errors</i>		<i>Adjusted <math>R^2</math></i>	
	<i>Pre-merger</i>	<i>Post-merger</i>	<i>Pre-merger</i>	<i>Post-merger</i>	<i>Pre-merger</i>	<i>Post-merger</i>
Infrastructure	0.6	0.5	0.25	0.25	0.29	0.31
Non-infrastructure	0.0	0.0	0.54	0.51	0.45	0.46
Management & general	1.2	2.1	0.23	0.23	0.25	0.22

Source: Ofwat, CC analysis.

32. Table 11 compares estimates of cost drivers’ parameters and their confidence intervals.

TABLE 11 Coefficients from Ofwat’s econometric models of capital maintenance expenditure pre- and post-merger

	<i>Point estimates of cost drivers’ coefficients</i>		<i>Confidence intervals around estimates</i>	
	<i>Pre-merger</i>	<i>Post-merger</i>	<i>Pre-merger</i>	<i>Post-merger</i>
Infrastructure	0.73	0.76	67.2	66.0
Non-infrastructure	0.83	0.80	49.2	49.2
Management & general	10.33	9.54	75.5	82.7

Source: Ofwat, CC analysis.

33. Table 12 compares Cook’s distances from Ofwat’s capital maintenance econometric models pre- and post-merger.

TABLE 12 Cook's distances from Ofwat's econometric models of capital maintenance expenditure pre- and post-merger

	Average		Maximum	
	Pre-merger	Post-merger	Pre-merger	Post-merger
Distribution infrastructure	0.03	0.04	0.14	0.14
Distribution non-infrastructure	0.07	0.08	0.49	0.53
Management and general	0.04	0.04	0.24	0.25

Source: Ofwat, CC analysis.

34. Table 12 shows that, on average, Cook's distances tend to increase as the number of comparators decreases. This means that post-merger, every observation has a greater effect on the results of the regression analysis than pre-merger. This follows naturally from the reduced number of observations post-merger. Table 12 does not indicate that there is an individual leverage point in any of the models, pre- or post-merger.

#### 4. Results of Ofwat's econometric models: changes over time versus changes through the merger

35. In this section we compare the variation in Ofwat's econometric models over time against the changes in the same models caused by the simulation of the merger.

##### (a) Operating expenditure

36. Table 13 lists (a) estimated coefficients, (b) width of the confidence intervals around these estimates, and (c) a measure of goodness of fit over a period of four years (2002/03 to 2005/06) for the operating expenditure model 'distribution'.

TABLE 13 Distribution operating expenditure model

Year	Point estimate of cost driver's coefficient		Confidence interval around estimated coefficient (%)		Proportion of variation in costs explained by variation in the cost driver (%)	
	Pre-merger	Post-merger	Pre-merger	Post-merger	Pre-merger	Post-merger
2002/03	13.27	36.75	123.5	62.3	0.36	0.79
2003/04	6.11	6.02	66.3	73.8	0.30	0.26
2004/05	5.01	4.96	75.2	82.9	0.24	0.21
2005/06	4.53	4.62	97.4	102.2	0.15	0.14

Source: Ofwat, CC analysis.

37. Table 13 illustrates that the variation between years exceeds the variation caused by the simulation of the merger in all years but 2002/03. For example, the estimated cost driver's coefficient, pre-merger, varies between 6.11 and 5.01 between 2003/04 and 2004/05. The same cost driver's coefficient varies between 6.11 and 6.02 pre- and post-merger in 2003/04.
38. Similarly, the width of the confidence interval varies between 66 and 75 per cent between 2003/04 and 2004/05, while it varies between 66 and 74 per cent pre- and post-merger in 2003/04.
39. Table 14 lists (a) estimated coefficients, (b) width of the confidence intervals around these estimates, and (c) a measure of goodness of fit over a period of four years (2002/03 to 2005/06).

TABLE 14 **Operating expenditure—resource and treatment**

Year	Point estimate of cost driver A's coefficient		Confidence interval around coefficient's point estimated (cost driver A), %		Point estimate of cost driver B's coefficient		Confidence interval around coefficient's point estimated (cost driver B), %	
	Pre-merger	Post-merger	Pre-merger	Post-merger	Pre-merger	Post-merger	Pre-merger	Post-merger
2002/03	16.77	16.06	78.2	86.8	5.12	5.12	100.0	102.7
2003/04	22.42	22.89	61.2	62.3	5.93	5.77	87.7	91.5
2004/05	18.48	19.39	82.4	81.5	4.13	3.99	138.6	145.8
2005/06	23.85	23.89	59.9	66.3	-5.18	-5.22	74.2	-78.6

Source: Ofwat, CC analysis.

40. Table 15 lists (a) estimated coefficients, (b) width of the confidence intervals around these estimates, and (c) a measure of goodness of fit over a period of four years (2002/03 to 2005/06) for the capital operating expenditure model 'power'.

TABLE 15 **Power operating expenditure model**

Year	Point estimate of cost driver's coefficient		Confidence interval around coefficient's point estimated %		Proportion of variation in costs explained by variation in the cost driver	
	Pre-merger	Post-merger	Pre-merger	Post-merger	Pre-merger	Post-merger
2002/03	0.94	0.94	5.0	5.0	0.99	0.99
2003/04	0.93	0.93	5.5	5.5	0.99	0.99
2004/05	0.93	0.93	6.2	6.2	0.98	0.98
2005/06	0.91	0.91	5.4	5.4	0.99	0.99

Source: Ofwat, CC analysis.

41. Table 16 lists (a) estimated coefficients, (b) width of the confidence intervals around these estimates, and (c) a measure of goodness of fit over a period of four years (2002/03 to 2005/06) for the capital operating expenditure model 'business activities'.

TABLE 16 **Business activities operating expenditure model**

Year	Point estimate of cost driver's coefficient		Confidence interval around coefficient's point estimated %		Proportion of variation in costs explained by variation in the cost driver	
	Pre-merger	Post-merger	Pre-merger	Post-merger	Pre-merger	Post-merger
2002/03	0.95	0.95	8.7	9.1	0.96	0.96
2003/04	0.92	0.92	9.5	9.8	0.96	0.96
2004/05	0.93	0.93	9.7	9.9	0.96	0.96
2005/06	0.95	0.95	7.3	7.5	0.98	0.98

Source: Ofwat, CC analysis.

### **(b) Capital maintenance expenditure**

42. Table 17 below lists (a) estimated coefficients, (b) width of the confidence intervals around these estimates, and (c) a measure of goodness of fit over a period of four years (2002/03 to 2005/06) for the capital maintenance expenditure model 'distribution infrastructure'.

TABLE 17 **Distribution infrastructure capital maintenance expenditure model**

Year	Point estimate of cost driver's coefficient		Confidence interval around coefficient's point estimated %		Proportion of variation in costs explained by variation in the cost driver	
	Pre-merger	Post-merger	Pre-merger	Post-merger	Pre-merger	Post-merger
2002/03	0.89	0.91	47.1	47.1	0.47	0.48

2003/04	0.76	0.79	61.6	61.7	0.33	0.34
2004/05	0.73	0.76	67.8	66.0	0.29	0.31
2005/06	0.74	0.78	69.0	66.9	0.28	0.31

Source: Ofwat, CC analysis.

43. Table 18 lists (a) estimated coefficients, (b) width of the confidence intervals around these estimates, and (c) a measure of goodness of fit over a period of four years (2002/03 to 2005/06) for the capital maintenance expenditure model 'distribution non-infrastructure'.

TABLE 18 Distribution non-infrastructure capital maintenance expenditure model

Year	Point estimate of cost driver's coefficient		Confidence interval around coefficient's point estimated %		Proportion of variation in costs explained by variation in the cost driver	
	Pre-merger	Post-merger	Pre-merger	Post-merger	Pre-merger	Post-merger
2002/03	0.66	0.63	65.2	65.1	0.31	0.32
2003/04	0.74	0.71	56.3	56.5	0.38	0.39
2004/05	0.83	0.80	49.2	49.2	0.45	0.46
2005/06	0.94	0.91	45.7	46.2	0.49	0.49

Source: Ofwat, CC analysis.

44. Table 19 lists (a) estimated coefficients, (b) width of the confidence intervals around these estimates, and (c) a measure of goodness of fit over a period of four years (2002/03 to 2005/06) for the capital maintenance expenditure model 'distribution non-infrastructure'.

TABLE 19 Management and general capital maintenance expenditure model

Year	Point estimate of cost driver's coefficient		Confidence interval around coefficient's point estimated, %		Proportion of variation in costs explained by variation in the cost driver	
	Pre-merger	Post-merger	Pre-merger	Post-merger	Pre-merger	Post-merger
2002/03	13.02	11.98	93.2	104.0	0.16	0.13
2003/04	12.77	12.09	90.1	98.1	0.17	0.15
2004/05	10.33	9.54	75.5	82.7	0.25	0.22
2005/06	9.17	9.44	75.9	76.3	0.25	0.26

Source: Ofwat, CC analysis.