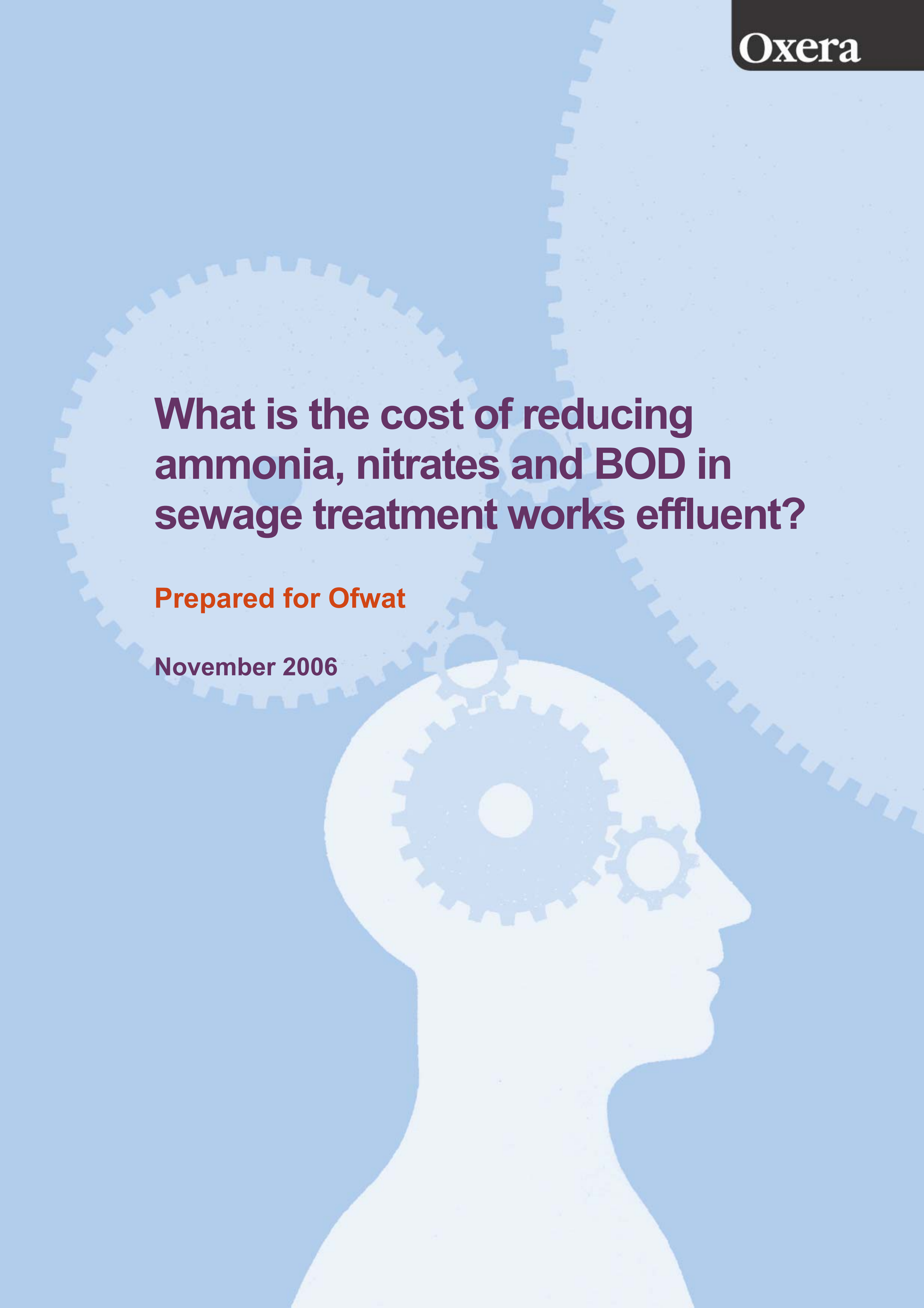
Oxera



What is the cost of reducing ammonia, nitrates and BOD in sewage treatment works effluent?

Prepared for Ofwat

November 2006

Oxera Consulting Ltd is registered in England No. 2589629. Registered office at Park Central, 40/41 Park End Street, Oxford OX1 1JD, UK.

Contents

List of tables

List of figures

1 Introduction

1.1 Background

The water industry may be required to play a key role where its activities prevent, or are forecast to prevent, achievement of Good Ecological Status under the Water Framework Directive (WFD), and where cost criteria (one of which is cost-effectiveness) are met for the work required. As part of the preparations, Defra is chairing a Collaborative Research Programme on the economic analysis required for implementation of the WFD. This includes a project to develop guidance and a methodology for identifying the most cost-effective measures for delivering environmental improvements. To compare the relative costs of measures taken by different industrial sectors, it is necessary to derive generic costs of carrying out measures that achieve environmental improvements.

1.2 Objective of research

Ofwat, together with Arup/Oxera, has previously undertaken research into the cost of reducing the level phosphate in sewage effluent.[1]

The objective of the present study is to obtain cost estimates for reducing the level of the following three parameters in sewage effluent:

- ammonia;
- nitrates;
- biochemical oxygen demand (BOD).

Ofwat has collated a dataset from the water companies on the estimated costs (operating expenditure, OPEX, and capital expenditure, CAPEX) of reducing the above parameters in sewage effluent, the associated policy variable (ie, the proposed consent), and other potential cost drivers. This dataset was provided to Oxera on an anonymised basis.

These cost estimates are to be used in an appraisal of the cost-effectiveness of delivering environmental improvements as part of the Collaborative Research Programme.

An element of caution should be applied when using the figures presented in this report since only a limited dataset has been used and many of the costs established are derived on the basis of various assumptions, which are explained throughout the report.

1.3 Structure of the report

This report summarises Oxera's main findings in analysing Ofwat's dataset and presents:

- econometric cost functions for each of the three parameters that represent the relationship between the equivalent annual cost (EAC) of reducing the three parameters in sewage effluent and the main cost drivers of these processes;

- unit cost estimates of treating the sewage effluent to certain standards (measured in £/kg of load removed).

---

[1] Ofwat, Arup and Oxera (2005),'Water Framework Directive: Economic Analysis of Water Industry Costs', November; and Oxera (2005), 'Review of Econometric Cost Modelling of Chemical Phosphorus Removal Works', report for Ofwat, November.

This report is structured as follows.

- Section 2 provides an overview of the Ofwat dataset that forms the basis of the analysis in this report.
- Section 3 describes the approach adopted in eliciting cost functions for the parameters and provides an overview of the main hypotheses tested.
- Section 4 presents cost functions for each parameter and briefly comments on the main findings of the econometric analysis.
- Section 5 provides unit cost look-up tables (£/kg parameter load removed) based on the econometric cost functions in section 4.
- Appendix 1 explains the different tertiary treatment types used at sewerage treatment works.
- Appendix 2 provides supplementary descriptive statistics.

2 Data

2.1 Description of dataset

For each parameter, Ofwat collated information at the observation level of individual sewage treatment works on costs (OPEX and CAPEX), cost drivers and the policy variable (proposed consent).

The Ofwat dataset includes the following variables.

Cost drivers

- Population equivalent;
- consented dry weather flow (cDWF, m^3/day);
- influent load (kg/day);
- effluent load (kg/day);
- load removed (kg/day);
- price index for capital costs (unit labour cost, Kw);
- price index for labour costs (construction price index, Lw);
- current treatment provision—this variable contains information on the type of treatment before implementing the changes at the sewage treatment work—ie, primary, secondary (activated sludge or biological filter) or tertiary (A1, A2, B1 or B2);
- proposed treatment provision—this variable contains information on the type of treatment after implementing the changes at the sewage treatment work—ie, primary, secondary (activated sludge or biological filter) or tertiary (A1, A2, B1 or B2).

Appendix 1 explains A1, A2, B1 and B2 in the context of tertiary treatment provision.

Policy variables

- Current consent—for BOD and nitrate in the BOD and nitrate dataset respectively, while in the ammonia dataset information on both ammonia and BOD levels is provided (milligrammes per litre, mg/l);
- proposed consent—for BOD and nitrate in the BOD and nitrate dataset respectively, while in the ammonia dataset information on both ammonia and BOD levels is provided (mg/l).

Total cost of proposed treatment

- Equivalent annual cost (EAC)—total CAPEX and OPEX expressed as a constant cost that arises each year over the assumed lifetime of the project, incurred in reducing each of the parameter levels;
- net present cost (NPC)— total CAPEX and OPEX incurred over the assumed lifetime of a project discounted to today's prices in reducing each of the parameter levels;
- unit costs (EAC)—total CAPEX and OPEX per annum expressed as a constant cost that arises each year over the assumed lifetime of the project, incurred in reducing each of the parameters per kg of load removed;
- unit costs (NPC)—total CAPEX and OPEX over the assumed lifetime of a project discounted to today's prices, incurred in reducing each of the parameters per kg of load removed.

Oxera has taken the cost associated with works as given. Ofwat's assumptions are that the civil engineering component of the investment would be 50% of CAPEX, the mechanical component 45%, and the electrical component 5%. The asset lives of these components were assumed to be 60, 20 and ten years respectively. Alternative assumptions could produce different total cost/unit cost levels. Additional discussion on the assumptions made in constructing the cost variable is provided in the previous research carried out by Ofwat, Arup and Oxera.[2]

In addition, using an anonymised sewage treatment works identifier variable contained in the dataset, Oxera constructed a company identifier variable. The identifier is a letter, the match of which is only known to Ofwat and individual companies.[3] The analysis is undertaken at the level of individual sewage treatment works (ie, the sites) rather than at the level of companies. Using the company identifier variable allows the identification of whether costs vary in a systematic manner by individual (albeit unknown) companies owing to (unknown) factors that are not reflected in the cost drivers and the policy variable (see further discussion in section 3).

In line with the previous work by Ofwat on phosphorus costs, the costs in this report are expressed in EAC.[4] All cost figures are expressed in financial year 2002/03 prices.

2.2 Data verification and descriptive statistics of dataset

2.2.1 Data verification

An essential element in ensuring the robustness of econometric analysis is the verification of the data used to model costs. Using statistical analysis, Oxera verified the plausibility of the information provided in the datasets and a list of anomalies was provided to, and discussed with, Ofwat. Where required, further clarification was sought from water companies that provided the data. Site observations with implausible values in any of the key variables were corrected or removed from the dataset. Table 2.1 shows the number of observations in each dataset as provided by Ofwat and following the verification process. In addition, the final column shows the number of observations that were included in the econometric models (section 4). The decision to exclude certain observations from the econometric analysis was based on statistical outlier tests.

Table 2.1 Number of sites for analysis

Dataset	Number of sites in dataset provided by Ofwat	Number of sites following data verification	Number of sites included in econometric modelling
Ammonia	440	387	355
Nitrates	95	95	84
BOD[1]	65	72	62

Note: [1] Each parameter was provided in a separate dataset. Seven sites were moved from the ammonia dataset to the BOD dataset due to the current and proposed ammonia consents being the same.
Source: Oxera.

2.2.2 Descriptive statistics

The strength of association between the variables in the dataset was examined using correlations. Correlation tables for the variables contained in each of the datasets are shown in Appendix 2. The tables provide the following key insights:

---

[2] Ofwat, Arup and Oxera (2005),'Water Framework Directive: Economic Analysis of Water Industry Costs', November; and Oxera (2005), 'Review of Econometric Cost Modelling of Chemical Phosphorus Removal Works', report for Ofwat, November.

[3] The letter does not allow third parties, including Oxera or companies other than the company concerned, to identify individual companies.

[4] While the data also contained separate information on CAPEX and OPEX, the analysis in this study focuses on total cost.

- **Correlation between EAC and measures of activity/scale**—there is a correlation of around 0.6 or above between costs and each of the following:

  - load removed;
  - population equivalent served;
  - cDWF;
  - influent load;
  - effluent load.

- **Correlation between EAC and proposed parameter consent**—a negative relationship between cost and the proposed consent was observed, so that a tighter proposed consent is associated with a higher cost, as might be expected. Contrary to what might be expected, for ammonia and BOD (see Tables A2.1 and A2.7), the correlation coefficients are low, suggesting a weak relationship. However, when using regression techniques that take into account the scale of works in addition to the consent and removing unusual observations, this relationship is strong and statistically significant (see section 4).

- **Correlations between measures of activity/scale**—a high correlation between individual measures of scale was observed (greater than 0.9 for ammonia and nitrates and 0.6 in BOD).

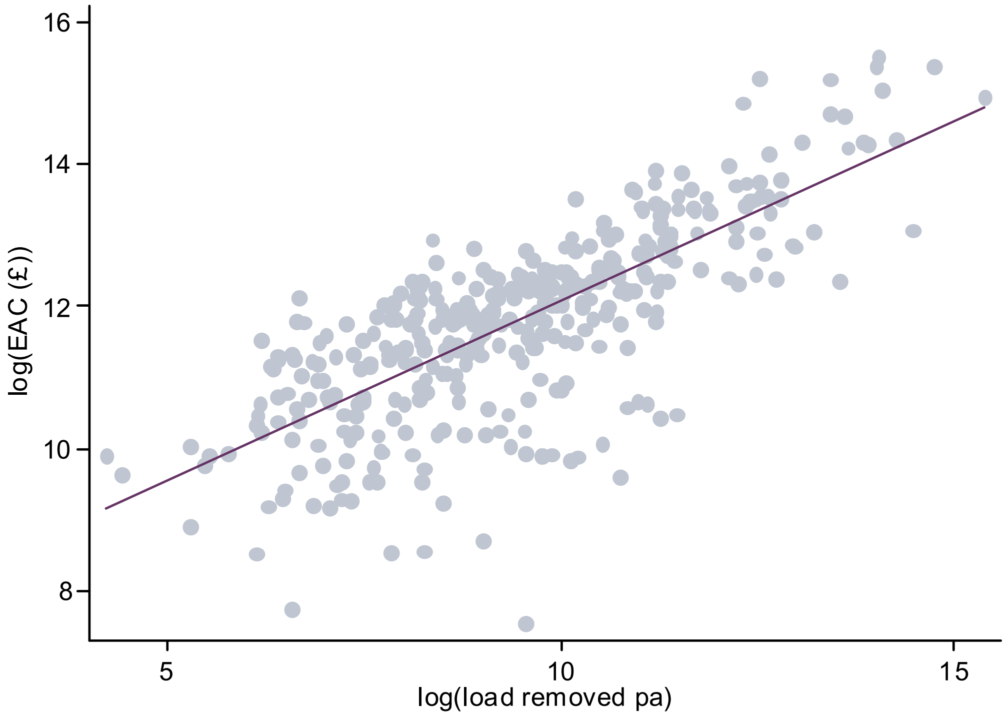
The high correlation between the measures of activity/scale suggests that each could be used individually as a measure of activity/scale in an econometric cost function as they capture similar processes. In line with the previous research on phosphorus costs, load removed is chosen as the key driver of costs at sewage treatment works.

Appendix 2 contains further descriptive statistics on the variables contained in the dataset.

2.2.3 Scatter plots

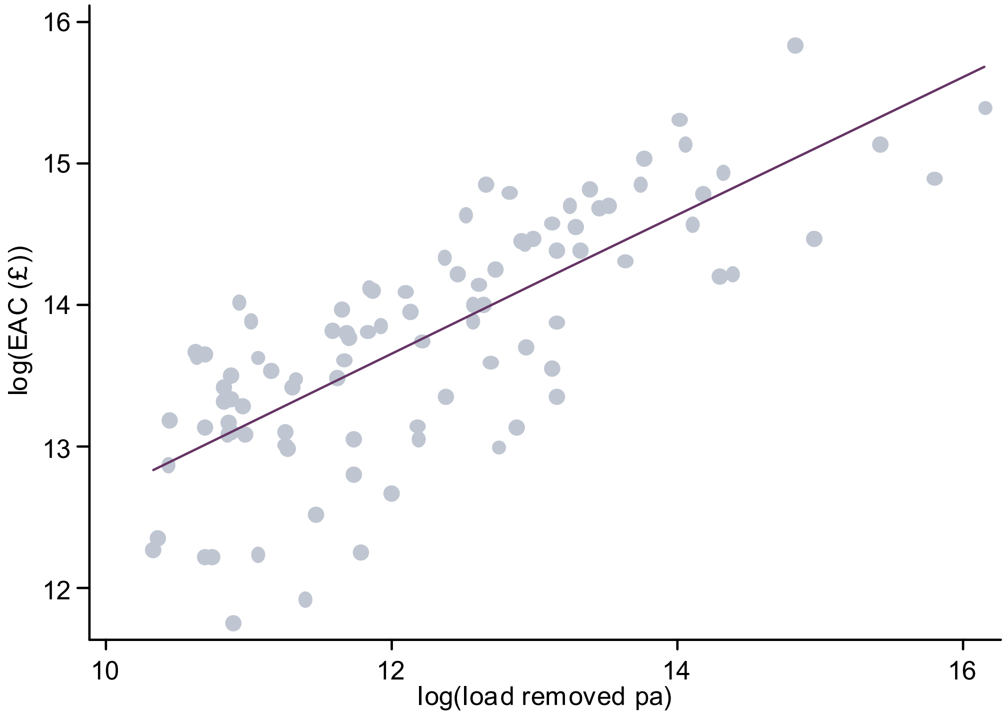
Plotting the cost–load relationship for each dataset, Figures 2.1 to 2.3 suggest that a broadly linear relationship (in logs) holds between cost and load removed. The figures highlight several potentially unusual observations—ie, high or low costs for a given level of load removed. The data is presented on a logarithmic scale since this allows the data that covers a large range of values to be effectively visualised. The line fitted to the scatter plot has a slope of less than 1. This indicates that the sewage treatment works are characterised by economies of scale, and hence that a doubling in scale leads to an increase in cost of less than twice the amount. This indicates that there are fixed costs associated with reducing the parameter, and hence unit costs fall as the size of works increases. The finding of economies of scale has also been tested and confirmed in the econometric models discussed in sections 3 and 4.

Figure 2.1 Ammonia: relationship between cost and load removed



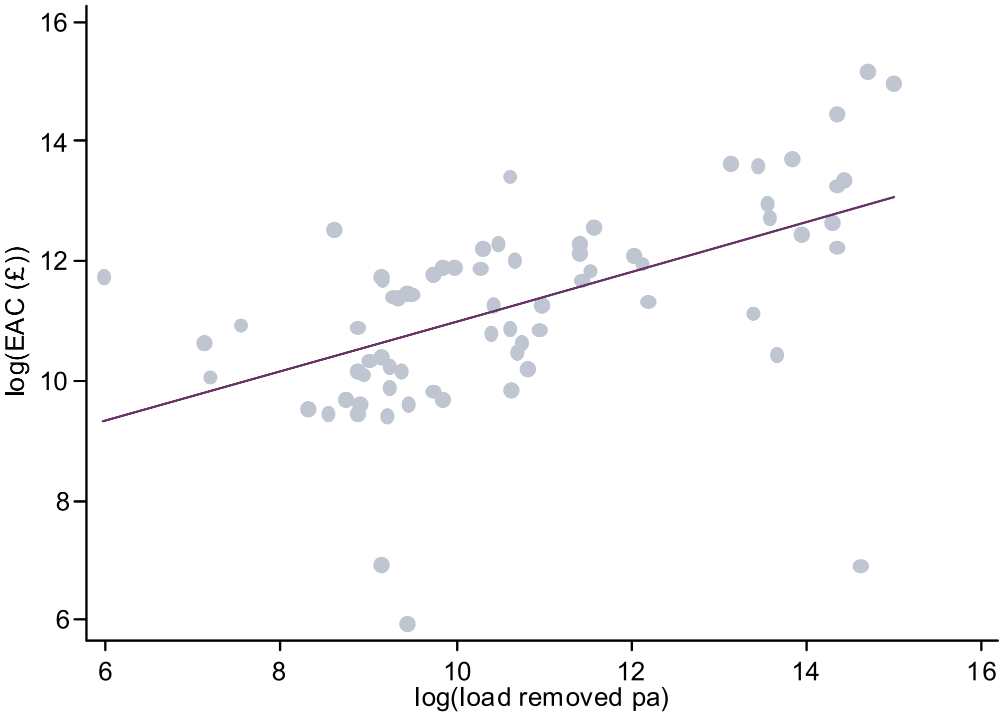
Source: Oxera.

Figure 2.2 Nitrates: relationship between cost and load removed



Source: Oxera.

Figure 2.3 BOD: relationship between cost and load removed



Source: Oxera.

Appendix 2 contains scatter plots of costs and proposed standards.

3 Approach, testing and hypotheses investigated

The approach to estimating the cost scenarios of reducing the parameters in sewage effluent involves the following two steps.

- Ordinary least squares (OLS) regression modelling of the sewage treatment work costs are used to estimate a cost function.[5] The modelling approach and overall insights gained are described in this section.

- The preferred cost functions obtained in the modelling exercise are provided in section 4. These are then used in section 5 to construct cost look-up tables to be used in cost-effectiveness analysis by Ofwat.

The cost functions are estimated using OLS techniques since the objective of is to identify a relationship between costs and cost drivers that reflects the *average* cost in the industry. Other techniques, such as corrected OLS, data envelopment analysis or stochastic frontier analysis as employed in the context of comparative efficiency measurement, allow the *minimum* cost to be identified for a given set of cost drivers. The unit cost estimates that could be identified using the information from these techniques would thus allow the minimum cost of reducing parameters from sewage effluent to be identified. These techniques could therefore be employed to identify a lower bound of the cost of reducing parameters in sewage effluent.

3.1 Modelling approach

A general-to-specific modelling methodology is employed to establish a cost function that summarises the relationship between costs and cost drivers. The approach examines a set of hypothesised relationships, beginning with a general model containing a range of candidate cost drivers. Using statistical hypotheses tests, the general model is simplified until a statistically preferred specification is reached. The identification of relevant hypotheses needs to be informed by economic and engineering principles.

In practice, the number of cost drivers that can be included in the general model is constrained by the limited cost driver data availability, and the size of the dataset (for BOD and nitrates). Within these constraints, the general-to-specific modelling exercise was undertaken. As highlighted above, due to the high correlation between the individual activity/scale measures, only a single measure of activity/scale could be included in a regression model at any one time along with other explanatory factors. In line with Ofwat's previous research on phosphorus costs, load removed was selected as the main cost driver.

3.2 Model testing

The general models were simplified or 'tested down' using statistical significance tests (t- and F-tests). The statistical appropriateness of the models was investigated and verified using the following standard diagnostic tests:

- normality of residuals;
- functional form;

---

[5] The OLS regression techniques allow quantification of the relationship that exists between two or more variables. The technique estimates coefficients in a linear model by minimising the sum of the squared differences between the observed dependent datapoints and those predicted by the linear regression model.

- heteroscedasticity.

Tables containing the results from the tests on the cost functions of each parameter are provided in section 4.

In addition, unusual observations (or ‘outliers’) may introduce bias in the estimated functional relationship and produce misleading cost predictions. Outlier detection tests are used to identify potentially influential observations in the econometric modelling. Observations identified as influential are investigated and removed or corrected, as appropriate, and the cost function is then re-estimated.

In addition, the unit of observation is the works level. Individual sites are not independent in that some common feature may mean that any given company can have costs that are systematically higher or lower than those of other companies. Assuming that no information is available on such company-specific factors, the underlying cause of the difference cannot be identified. Alternatively, it is possible that a company employs a different costing/cost allocation process and, as such, the site observations of a company are not directly comparable with those of other companies.

As a result, it may be possible that the industry (or the sample of works observed) is not characterised by a common cost function, and if such systematic (unobserved) differences exist, the average cost of works would be under- or overestimated for such companies.

This possibility was investigated by testing whether indicators identifying individual (but anonymous) companies are statistically significant in the cost function. The results for each parameter are presented in section 4.

3.3 Hypotheses investigated

Table 3.1 summarises the main hypotheses tested as part of the general-to-specific modelling approach for all three parameters. The tests undertaken that are specific to individual models are discussed in section 4, and the findings for each parameter are presented.

Table 3.1 Hypotheses tested in all cost functions

Hypothesis	Form of hypothesis test	Result of statistical test
Works are characterised by economies of scale	Test whether there are constant returns to scale (test of whether coefficient on load removed is equal to 1)	Constant returns to scale is rejected. Works are characterised by economies of scale
Small sites are governed by a different cost function than larger sites	Allow for an interaction between size (using indicator variables for different size bands) and load removed	Non-linearity of cost function is rejected. The constant relationship over the entire range confirms the observation of a broadly (log) linear relationship in scatter plots in section 2
The method of proposed treatment provision is an important cost driver	Using indicator variables for different treatment provisions, investigate whether costs are significantly different	Costs are statistically not systematically higher for different treatment types. While, in practice, the cost drivers are likely to be important, the data analysed does not allow discrimination between different types of proposed treatment
The method of proposed treatment provision systematically varies with the tightness of the proposed consent	Using indicator variables for different treatment provisions and tightness of consent, investigate whether costs are significantly different	Treatment provision does not vary systematically with the tightness of consent. This is in contrast with general industry experience and may be due to data limitations
The change in the consent is an important cost driver	Investigate whether the change between current and proposed consent is important as: a) as an additional cost driver, or b) as a main cost driver (instead of load removed)	a) Various model specifications in which the change in consent is included in addition to other cost drivers show that the change in consent is not a statistically significant cost driver b) Explanatory power of cost function is significantly less than when using load removed. This is in contrast with industry experience according to which the change in consent and load removed are seen as very similar
Company-wide differences in construction costs (Kw) and labour costs (Lw) are important cost drivers. (Factor prices are not available at the regional level)	Using variables capturing the construction cost (Kw) and labour cost (Lw), investigate whether costs are significant cost drivers	Company-wide differences do not seem to plausibly capture systematic differences in cost. While in some specifications Kw and Lw are statistically significant, the direction of cost impact varies across models. Due to economically counterintuitive results, factor price indicators were not used in the econometric analysis

Note: The types of proposed treatment provision in the dataset are secondary treatment activated sludge; secondary treatment biological; tertiary treatment A1; tertiary treatment A2; tertiary treatment B1; and tertiary treatment B2. See Appendix 1 for a description of different types of tertiary treatment.
Source: Oxera.

3.4 Preferred cost function specification

Based on the statistical testing of hypotheses and diagnostic testing, the following log linear cost function was identified as the preferred general specification:

$$\text{Log(EAC)} = a + b_1\text{log(load removed)} + b_2\text{log(proposed parameter standard)}$$

or:

$$\text{Log(EAC per kg of load removed)} = a + b_1\text{log(load removed)} + b_2\text{log(proposed parameter standard)}$$

where: a is a constant term; and b_1 and b_2 are the estimated coefficients of the cost function, which can be interpreted as elasticities.

Section 4 presents further details of each parameter modelled.

4 Modelling summary: results and commentary

4.1 Modelling results: ammonia

Table 4.1 shows the statistically preferred specification of the cost function for reducing ammonia in effluent.

The coefficients in a log linear model (ie, b_1 and b_2 in the equations in section 3) can be interpreted as elasticities. Table 4.1 shows that, on average, a 10% increase in load removed leads to a 5% increase in EAC. Similarly a 10% higher proposed ammonia consent is associated with a reduction of around 4% in EAC.

Table 4.1 Ammonia cost function

	Parameter	Estimated parameter of log linear cost function	t-statistics
Dependent variable		log (EAC per annum)	
Log (load removed per annum)	b_1	0.497	(26.29)**
Log (proposed ammonia consent)	b_2	–0.374	(5.29)**
Constant	a	7.729	(33.65)**
Observations		355	
R-squared		0.71	

Note: Absolute value of t-statistics in parentheses; * significant at 5%; ** significant at 1%.
Source: Oxera.

In addition to the hypotheses described in section 3.3, the following was investigated.

- **Joint reduction in ammonia and BOD**—a number of works (around 40% of the sites in the sample) undergo a reduction in both ammonia and BOD. A test of whether the removal of both ammonia and BOD increases costs over and above those of works that reduce only ammonia showed that the costs in the two types of works are not statistically significantly different. This is consistent with the fact that, in many instances, the treatment installed to remove either ammonia or BOD will act to remove the other parameter at the same time. (In the modelling the coefficient was positive yet statistically insignificant, which may indicate higher costs at some sites where both ammonia and BOD are removed.)

- **Systematic differences in costs between companies**—the inclusion of company indicators in the cost function reveals that, in the current dataset, companies *v* and *x* have significantly lower costs than the rest of the companies. The finding of a significantly lower cost could be due to companies operating in a different, lower-cost environment. If the lower cost is due to the existence of special factors and other unobserved variables, excluding the sites from the sample would be warranted. An average cost function that does not exclude the lower-cost companies would systematically under-predict costs for the industry as a whole and, as such, may lead to an underestimation of the representative industry cost. However, if these companies are reflective of the 'true' cost of reducing ammonia in sewage effluent, and other companies are too expensive, these observations should be included in the analysis and perhaps greater weight should be given to the observations of these companies. Since it is not known which of these possibilities applies, the unit cost scenarios in section 5 indicate the extent to which costs are higher when the fact that companies *v* and *x* have lower costs is taken into account in the modelling (via indicator variables).

The cost function presented in Table 4.1 passes the standard diagnostic tests for model misspecification. In addition, outlier tests have identified and excluded from the analysis influential observations that may bias the cost function parameter estimates. The results from the statistical tests are summarised in Table A2.11 in Appendix 2.

4.2 Modelling results: nitrates

Table 4.2 shows the statistically preferred specification of the cost function for reducing nitrates in effluent, with coefficients representing elasticities. The table shows that a 10% increase in load removed leads to a 5% increase in EAC.

Table 4.2 Nitrates cost function

	Parameter	Estimated parameter of log linear cost function	t-statistics
Dependent variable		log (EAC per annum)	
Log (load removed per annum)	b_1	0.466	(11.24)**
Constant	a	8.068	(15.62)**
Observations		84	
R-squared		0.60	

Note: Absolute value of t-statistics in parentheses; * significant at 5%; ** significant at 1%.
Source: Oxera.

In addition to the hypotheses described in section 3.3, the following was investigated.

- **Proposed nitrate consent**—the dataset includes only two proposed consents for nitrates. Statistical testing show that, for a given level of load removed, costs are not significantly higher for a consent of 10mg/l compared with a consent of 15mg/l. Given the statistical significance of the proposed consent in the other cost functions examined in this report, in principle the proposed consent could also be a significant cost driver if a wider range of consents were observed.

- **Systematic differences in costs between companies**—the inclusion of company indicators in the cost function reveals that, in the current dataset, companies *v* and *y* have significantly higher costs than the rest of the works. At the same time these two companies do not have any tertiary A2 or B2 treatment at any of their works (see Figure A2.5). Tertiary A2 and B2 treatment are often regarded as more resource-intensive. The finding that works *without* these types of treatment have higher costs is thus counterintuitive and it is likely that the cost differential is driven by different (unknown) factors.

Around two-thirds of the sample of works are from a single company (company *c*, see Figure A2.5 in Appendix 2). Moreover, this company employs a single type of treatment (tertiary A2). As a result, the estimated functional relationship is driven by company *c.* It therefore may not be possible to apply this cost function to the industry as a whole. This is particularly the case if company *c* has some unique feature across its sites. Any conclusions regarding the average cost of reducing nitrates based on a sample of sites dominated by a single company should therefore be made with due caution.

The cost function presented in Table 4.2 passes the standard diagnostic tests for model misspecification. In addition, outlier tests identified and excluded from the analysis influential observations that may bias the cost function parameter estimates. The results from the statistical tests are summarised in Table A2.12 in Appendix 2.

4.3 Modelling results: BOD

Table 4.3 shows the statistically preferred specification of the cost function for reducing BOD in effluent, with coefficients representing elasticities. The table shows that a 10% increase in load removed leads to a 4% increase in EAC. Similarly a 10% higher proposed BOD consent is associated with a 6% reduction in EAC.

Table 4.3 BOD cost function

	Parameter	Estimated parameter of log-linear cost function	t-statistics
Dependent variable		log (EAC per annum)	
Log (load removed per annum)	b_1	0.401	(6.96)**
Log (proposed BOD consent)	b_2	–0.581	(2.83)*
Constant	a	8.660	(8.69)**
Observations		62	
R-squared		0.59	

Note: Absolute value of t-statistics in parentheses; * significant at 5%; ** significant at 1%.
Source: Oxera.

In addition to the hypotheses described in section 3.3, the following was investigated.

– **Systematic differences in costs between companies**—the inclusion of company indicators in the cost function reveals that, in the current dataset, companies *c* and *z* have significantly higher costs than the rest of the industry. As such, using the above-average cost function to predict costs for the industry as a whole leads to an overestimation of the representative cost. The finding of a significantly higher cost could be due to companies operating in a different, higher-cost environment. If the higher cost were due to the existence of special factors and other unobserved variables, exclusion of the sites from the sample would be warranted to obtain an estimate of the industry cost excluding the potentially unusual observations. An average cost function that does not exclude the higher-cost companies would systematically over-predict costs for the industry as a whole and, as such, may lead to an overestimation of the representative industry cost. However, if these companies are reflective of the 'true' cost, and other companies are systematically underestimating their cost, then these observations should be included in the analysis. The existence of actual or apparent differences in cost between companies, together with a small sample size mean that the unit costs based on the above model should be interpreted and used with caution.

The cost function presented in Table 4.3 passes the standard diagnostic tests for model misspecification. In addition, outlier tests have been performed to identify and exclude from the analysis influential observations that may bias the cost function parameter estimates. The results from the statistical tests are summarised in Table A2.13 in Appendix 2.

5 Unit cost scenarios

This section presents the cost scenarios for each of the parameters.

The preferred cost functions from the econometric modelling exercise as outlined in section 4 were used to construct unit cost scenarios for the removal of the parameters following the template set out by Ofwat for phosphorus costs.[6]

To construct the cost scenario tables, additional assumptions are required. First, assumptions regarding the appropriate size bandings have been made. The ammonia bandings were requested by companies and Ofwat assisted in identifying appropriate size bands for nitrates and BOD. For a high level of confidence in estimates, it is important that a sufficient number of observations are contained in the dataset. As an indicator of the size in each size band the arithmetic average of the cost driver (ie load removed) was used. Table 5.1 shows the size bandings and the number of observations.

Table 5.1 Assumptions on size bandings and number of sites in each band

Ammonia	Number of sites	Nitrates	Number of sites	BOD	Number of sites
pe<500	34	**8,000<pe<100,000**	53	**<500**	17
500<=pe<2,000	76	**pe>100,000**	31	**500<=pe<4,000**	26
2,000<=pe<10,000	119			**4,000<=pe<50,000**	14
10,000<=pe<40,000	73			**pe>=50,000**	5
40,000<=pe<80,000	14				
80,000<=pe<150,000	22				
pe>=150,000	17				
Total number of observations	355		84		62

Note: pe, population equivalent.
Sources: Water companies and Ofwat.

Second, the tightness of consent needs to be broadly representative of the range of consents. It is therefore necessary for a sufficient number of observations to be included in each size band in order to have confidence in the cost prediction. Since there are only five works with a population equivalent of more than 50,000 in the BOD sample, caution should be exercised in making inferences based on such a small sample.

Tables 5.2 to 5.4 contain unit cost scenarios for each of the parameters. The confidence intervals shown refer to the average company observation of a given size in the *sample examined*. They therefore do not provide an indication of the range within which a particular sewage treatment works from the current, or indeed future, *population* of works from outside the observed sample is likely to lie.

To the extent that the current sample is broadly representative of the population of works (in terms of size, consent levels and other characteristics) that are likely to be built in the future, the confidence ranges represent a reasonable upper and lower bound of the expected costs.

---

[6] Ofwat, Arup and Oxera (2005), 'Water Framework Directive: Economic Analysis of Water Industry Costs', November.

However, if the structure of the industry is likely to be considerably different (in terms of size, consent levels and other characteristics), the construction of the confidence intervals needs to take into account the additional source of uncertainty from factors that have not been observed but that also have an impact on the cost of works (because data is not present in the sample). When making predictions about costs outside the sample, the confidence intervals therefore would become significantly wider.

Table 5.2 shows the unit costs for reducing ammonia for a proposed consent of 3mg/l, 5mg/l and 10mg/l. As discussed above, companies *v* and *x* have statistically significantly lower costs than other companies. Allowing for the systematic difference in cost increases the average cost for companies other than companies *v* and *x* by 8%, 6% and 3% for the 3mg/l, 5mg/l and 10mg/l proposed consents respectively.

Table 5.2 Ammonia: unit EAC per kg of load removed (£/kg)

	3mg/l			5 mg/l			10 mg/l		
Works size	**2.5% lower bound**	**Central estimate**	**2.5% upper bound**	**2.5% lower bound**	**Central estimate**	**2.5% upper bound**	**2.5% lower bound**	**Central estimate**	**2.5% upper bound**
pe<500	76.2	88.6	103.0	64.1	73.1	83.4	48.0	56.4	66.3
500<=pe<2,000	38.6	43.1	48.0	32.4	35.6	39.0	23.8	27.4	31.6
2,000<=pe<10,000	18.9	20.5	22.2	15.7	16.9	18.2	11.3	13.1	15.0
10,000<=pe<40,000	8.8	9.6	10.5	7.2	7.9	8.7	5.2	6.1	7.2
40,000<=pe<80,000	5.5	6.1	6.8	4.5	5.1	5.7	3.3	3.9	4.7
80,000<=pe<150,000	3.6	4.1	4.6	2.9	3.3	3.8	2.1	2.6	3.2
pe>=150,000	1.8	2.1	2.5	1.4	1.7	2.1	1.1	1.3	1.7

Note: pe, population equivalent. For each consent level the central column presents the estimated value with the left- and right-hand columns being the lower and upper bounds of the 95% confidence interval respectively.
Source: Oxera.

Table 5.3 shows the estimated unit costs for reducing nitrates in sewage effluent for two average works sizes. As discussed above, companies *v* and *x* have statistically significantly higher costs than other companies. In addition, two-thirds of the sample are from a single company, so that it is possible that the observed sample is not representative of the industry as a whole. This may need to be considered in further use of these figures. Allowing for these systematic differences in cost reduces the average costs for companies, other than companies *v* and *y*, by 26% and 23% for the 10mg/l and 15mg/l proposed consents respectively. If only company *c*'s data is analysed, costs are on average around 8% higher for the size banding with smaller works, but almost 30% higher for the larger works.

Table 5.3 Nitrates: unit EAC per kg of load removed (£/kg)

Works size	**2.5% lower bound**	**Central estimate**	**2.5% upper bound**
8000<pe<100,000	6.5	7.4	8.4
pe>100,000	1.9	2.2	2.6

Note: pe, population equivalent. The central column presents the estimated value, with the left- and right-hand columns being the lower and upper bounds of the 95% confidence interval respectively.
Source: Oxera.

Table 5.4 shows the unit costs for reducing BOD for a proposed consent of 10mg/l and 15mg/l. As discussed above, companies *c* and *z* have statistically significantly higher costs than the rest of companies. Allowing for these systematic differences in cost reduces the average costs for companies, other than companies *c* and *z*, on average by 26% and 23% for the 10mg/l and 15 mg/l proposed consents respectively.

Table 5.4 BOD: unit EAC per kg of load removed (£/kg)

	10 mg/l			15 mg/l		
Works size	**2.5% lower bound**	**Central estimate**	**2.5% upper bound**	**2.5% lower bound**	**Central estimate**	**2.5% upper bound**
<500	8.7	13.0	19.4	7.6	10.3	13.9
500<=pe<4,000	4.1	5.5	7.5	3.5	4.4	5.4
4,000<=pe<50,000	1.0	1.3	1.8	0.9	1.1	1.3
Pe>=50,000	0.3	0.5	0.7	0.2	0.4	0.6

Note: pe, population equivalent. For each consent level the central column presents the estimated value, with the left- and right-hand columns being the lower and upper bounds of the 95% confidence interval respectively.
Source: Oxera.

Appendix 1 Explanation of the tertiary treatments

Ofwat distinguishes between works with tertiary treatment using the following four categories.

- **Tertiary A1**—works with a secondary activated sludge process, the treatment methods of which also include prolonged settlement in conventional lagoons or raft lagoons, irrigation over grassland, constructed wetlands, root zone treatment (where used as a tertiary stage), drum filters, microstrainers, slow sand filters, tertiary nitrifying filters, wedge wire clarifiers, or Clariflow installed in humus tanks (where used as a tertiary treatment stage).

- **Tertiary A2**—works with a secondary activated sludge process, the treatment methods of which also include rapid-gravity sand filters, moving bed filters, pressure filters, nutrient control using physio-chemical and biological methods, disinfection, hard COD and colour removal, where used as a tertiary treatment stage.

- **Tertiary B1**—works with a secondary stage biological process, the treatment methods of which also include prolonged settlement in conventional lagoons or raft lagoons, irrigation over grassland, constructed wetlands, root zone treatment (where used as a tertiary stage), drum filters, microstrainers, slow sand filters, tertiary nitrifying filters, wedge wire clarifiers or Clariflow installed in humus tanks (where used as a tertiary treatment stage).

- **Tertiary B2**—works with a secondary biological process, the treatment methods of which also include rapid gravity sand filters, moving bed filters, pressure filters, nutrient control using physio-chemical and biological methods, disinfection, hard COD and colour removal, where used as a tertiary treatment stage.[7]

---

[7] Source: Ofwat.

Appendix 2 Additional descriptive statistics

A2.1 Descriptive statistics

Table A2.1 Ammonia dataset correlations

	EAC	Unit EAC	Load removed	PE	cDWF	Influent	Effluent	Amm current	Amm proposed	BOD current	BOD proposed	Kw	Lw
EAC	1.000												
Unit EAC	–0.168	1.000											
Load removed	0.673	–0.160	1.000										
PE	0.677	–0.163	1.000	1.000									
cDWF	0.703	–0.168	0.959	0.963	1.000								
Influent	0.677	–0.163	1.000	1.000	0.963	1.000							
Effluent	0.694	–0.190	0.916	0.926	0.962	0.926	1.000						
Amm current	–0.127	0.297	–0.115	–0.115	–0.139	–0.115	–0.115	1.000					
Amm proposed	–0.121	0.297	–0.082	–0.075	–0.084	–0.075	0.032	0.363	1.000				
BOD current	0.031	0.160	–0.027	–0.022	–0.027	–0.022	0.050	0.261	0.350	1.000			
BOD proposed	0.024	0.190	–0.030	–0.024	–0.024	–0.024	0.062	0.202	0.504	0.767	1.000		
Kw	–0.162	–0.014	–0.080	–0.082	–0.070	–0.082	–0.110	0.161	–0.122	–0.150	–0.245	1.000	
Lw	0.002	–0.021	–0.020	–0.020	–0.026	–0.020	–0.021	0.055	0.105	0.117	0.049	–0.029	1.000

Source: Oxera.

Table A2.2 Ammonia dataset descriptive statistics

Statistic	EAC (£)	Unit EAC (£/kg load removed)	Load removed (kg/day)	PE	cDWF (m^3/day)	Influent (kg/day)	Effluent (kg/day)	Amm current (mg/l)	Amm proposed (mg/l)	BOD current (mg/l)	BOD proposed (mg/l)
mean	320,000	22.96	255.41	39.47	10,099.71	276.31	20.89	23.85	5.22	25.23	19.68
p25	58,788	4.42	8.29	1.46	339.00	10.24	0.91	6.00	3.00	15.00	11.00
p50	140,000	11.19	33.24	5.24	1,364.50	36.68	2.75	10.00	3.00	20.00	15.00
p75	270,000	25.07	126.18	19.30	5,200.00	135.09	10.53	26.00	5.00	30.00	25.00
min	1,892	0.14	0.19	0.04	10.00	0.25	0.00	2.00	1.00	0.00	4.00
max	5,400,000	295.38	13,142.59	1,983.58	450,000.00	13,885.09	742.50	70.00	40.00	400.00	150.00
sd	640,000	34.52	933.33	141.51	34,612.66	990.57	62.13	25.51	4.86	23.59	13.04
N	387	387	387	387	386	387	387	385	387	386	387

Source: Oxera.

Table A2.3 Ammonia dataset EAC (£) per kg of load removed by size band

Size (population equivalent)	Mean	25th percentile	50th percentile	75th percentile	Minimum	Maximum	Standard deviation	N
pe<500	84.8	41.1	73.4	116.9	3.3	295.4	64	42
500<=pe<2,000	34.3	13.5	26.9	53.1	1.4	132.5	26.2	81
2,000<=pe<10,000	14.3	7.7	13.2	18.5	0.1	51.5	9.5	127
10,000<=pe<40,000	6.5	3.9	5.8	8.2	0.3	28.5	4.6	81
40,000<=pe<80,000	5.1	2.8	5.2	6.5	1.3	10.6	2.8	15
80,000<=pe<150,000	3.7	2.1	2.8	3.5	1	15.1	3.5	22
pe>=150,000	2.1	0.9	1.6	3.6	0.2	6	1.6	19
Overall sample	23	4.4	11.2	25.1	0.1	295.4	34.5	387

Source: Oxera.

Table A2.4 Nitrates dataset correlations

	EAC	Unit EAC	Load removed	PE	cDWF	Influent	Effluent	Nitrates proposed	Kw	Lw
EAC	1.000									
Unit EAC	–0.405	1.000								
Load removed	0.583	–0.459	1.000							
PE	0.583	–0.458	1.000	1.000						
cDWF	0.514	–0.394	0.974	0.974	1.000					
Influent	0.583	–0.458	1.000	1.000	0.974	1.000				
Effluent	0.508	–0.395	0.971	0.971	0.999	0.971	1.000			
Nitrates proposed	–0.702	0.579	–0.563	–0.563	–0.482	–0.563	–0.459	1.000		
Kw	0.111	–0.273	0.639	0.639	0.643	0.639	0.641	–0.295	1.000	
Lw	–0.111	0.273	–0.639	–0.639	–0.643	–0.639	–0.641	0.295	–0.413	1.000

Source: Oxera.

Table A2.5 Nitrates dataset descriptive statistics

Statistic	EAC (£)	Unit EAC (£/kg load removed)	Load removed (kg/day)	PE	cDWF (m^3/day)	Influent (kg/day)	Effluent (kg/day)	Nitrates proposed (mg/l)
mean	1,400,000	6.28	1,680.89	162.25	1,596.39	1,711.71	30.82	13.26
p25	520,000	2.69	174.66	19.11	206.01	201.65	0.08	10.00
p50	990,000	4.68	505.02	47.88	540.62	505.16	0.20	15.00
p75	1,900,000	8.71	1,403.02	134.43	1,480.80	1,418.18	3.99	15.00
min	130,000	0.41	83.76	8.78	92.58	92.61	0.02	10.00
max	7,500,000	22.29	28,478.81	2,700.00	19,935.51	28,485.00	869.00	15.00
sd	1,200,000	4.90	3,904.59	370.56	3,077.03	3,909.43	100.25	2.39
N	95	95	95	95	72	95	95	95

Note: Only 72 sites provided plausible values for cDWF.
Source: Oxera.

Table A2.6 Nitrates dataset EAC (£) per kg of load removed by size band

Size (population equivalent)	Mean	25th percentile	50th percentile	75th percentile	Minimum	Maximum	Standard deviation	N
8,000<=pe<100,000	8.2	4.5	7.3	10.3	22.3	22.3	5.8	4.9
pe>=100,000	2.5	1.3	2.5	3.6	4.8	4.8	2.3	1.4
Overall sample	6.3	2.7	4.7	8.7	22.3	22.3	6	4.9

Source: Oxera.

Table A2.7 BOD dataset correlations

	EAC	Unit EAC	Load removed	PE	cDWF	Influent	Effluent	BOD proposed	Kw	Lw
EAC	1.000									
Unit EAC	–0.044	1.000								
Load removed	0.735	–0.111	1.000							
PE	0.738	–0.111	1.000	1.000						
cDWF	0.689	–0.070	0.673	0.684	1.000					
Influent	0.738	–0.111	1.000	1.000	0.684	1.000				
Effluent	0.617	–0.075	0.648	0.661	0.951	0.661	1.000			
Nitrates proposed	–0.210	–0.082	–0.135	–0.136	–0.170	–0.136	–0.110	1.000		
Kw	–0.243	0.049	–0.249	–0.249	–0.200	–0.249	–0.146	0.758	1.000	
Lw	0.271	–0.117	0.252	0.253	0.238	0.253	0.198	–0.238	–0.382	1.000

Source: Oxera.

Table A2.8 BOD dataset descriptive statistics

Statistic	EAC (£)	Unit EAC (£/kg load removed)	Load removed (kg/day)	PE	cDWF (m^3/day)	Influent (kg/day)	Effluent (kg/day)	BOD current (mg/l)	BOD proposed (mg/l)
mean	270,000	9.19	939.07	15,914.87	2,075.84	954.89	15.83	36.08	19.01
p25	26,612	0.88	26.61	471.50	52.25	28.29	0.53	20.00	10.00
p50	89,163	1.87	84.65	1,508.00	159.50	90.48	1.73	35.00	15.00
p75	200,000	6.55	479.25	8,169.50	1,000.00	490.17	8.37	40.00	25.00
min	379	0.00	1.08	49.00	0.20	2.94	0.00	10.00	5.00
max	3,900,000	325.00	8,884.76	150,000.00	30,000.00	8,970.63	241.89	110.00	43.00
sd	630,000	38.62	1,863.19	31,512.92	5,487.66	1,890.78	42.12	19.58	10.58
N	72	72	72	72	72	72	72	72	72

Note: Only 72 sites provided plausible values for cDWF.
Source: Oxera.

Table A2.9 BOD dataset EAC (£) per kg of load removed by size band

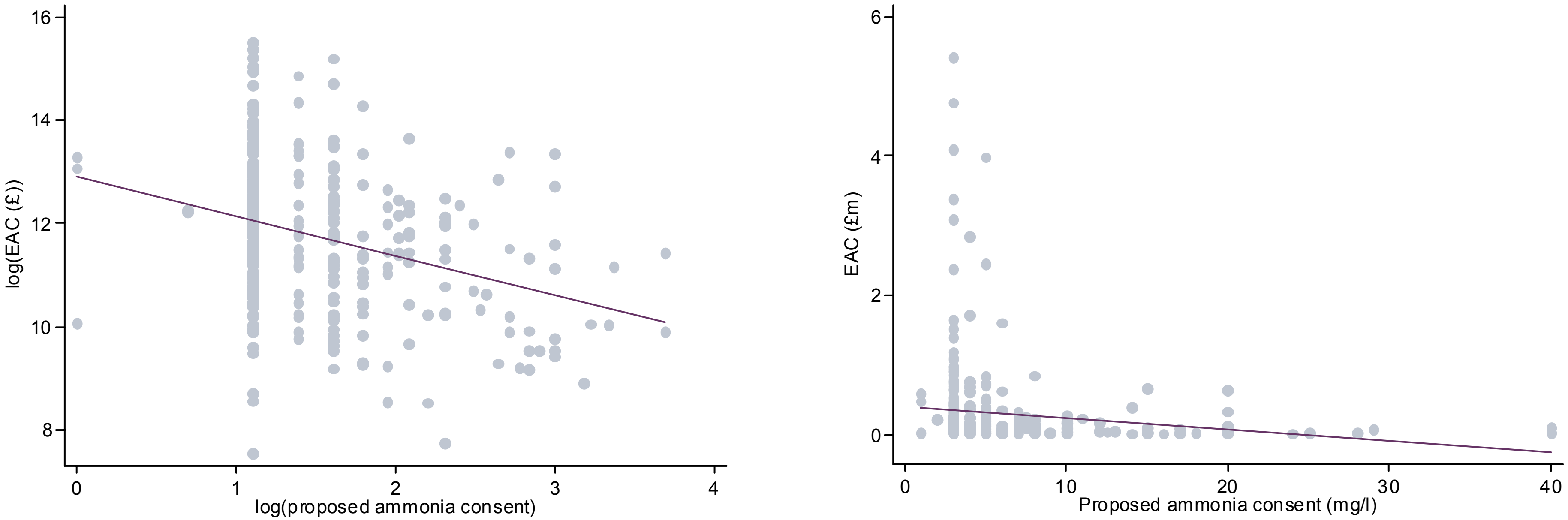
Size (population equivalent)	Mean	25th percentile	50th percentile	75th percentile	Minimum	Maximum	Standard deviation	N
<500	26.50	2.70	3.80	15.70	0.10	325.00	71.50	20.00
500<=pe<4,000	3.90	1.00	2.10	7.10	0.00	16.90	3.90	28.00
4,000<=pe<50,000	1.20	0.40	1.10	1.70	0.00	2.80	0.80	15.00
Pe>=50,000	0.60	0.20	0.30	1.00	0.00	1.60	0.60	9.00
Overall sample	9.20	0.90	1.90	6.50	0.00	325.00	38.60	72.00

Source: Oxera.

A2.2 Relationship between cost and proposed consent

A logarithmic scale is used since this allows data with a large range of values to be presented more effectively than when presented in levels. In addition, for ammonia and BOD, the same information is also displayed in levels.

Figure A2.1 Ammonia: relationship between cost and proposed consent

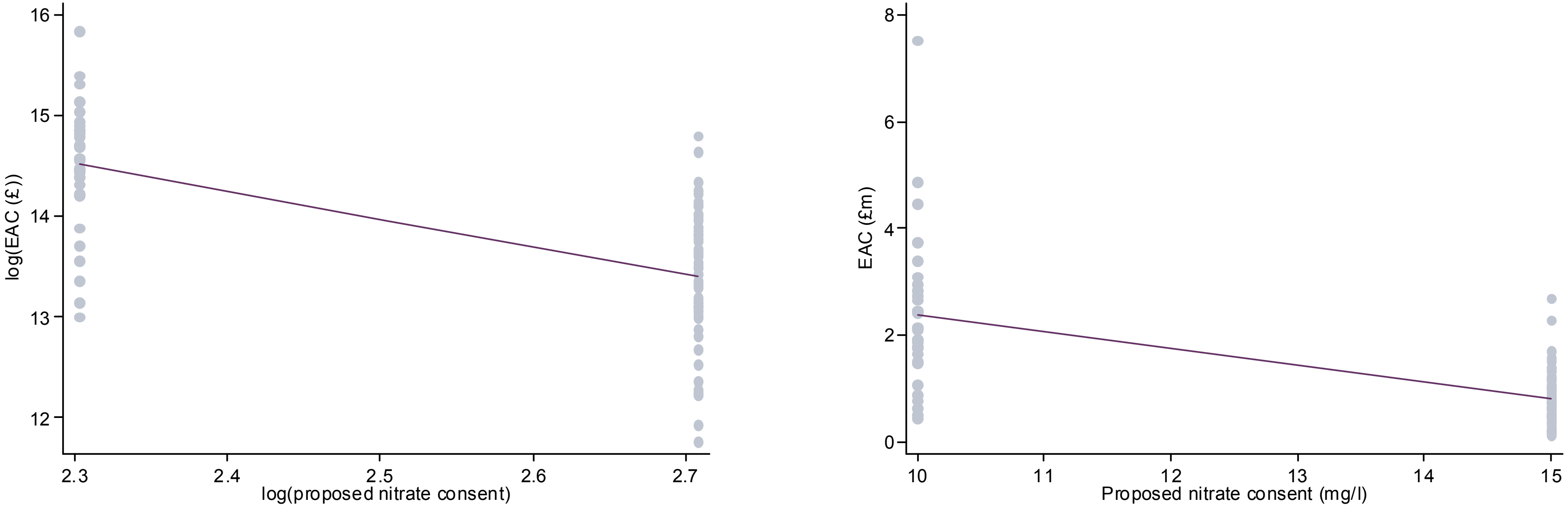


Scatter plot of logarithms of costs and proposed consent

Scatter plot of costs and proposed consent

Source: Oxera.

Figure A2.2 Nitrates: relationship between cost and proposed consent

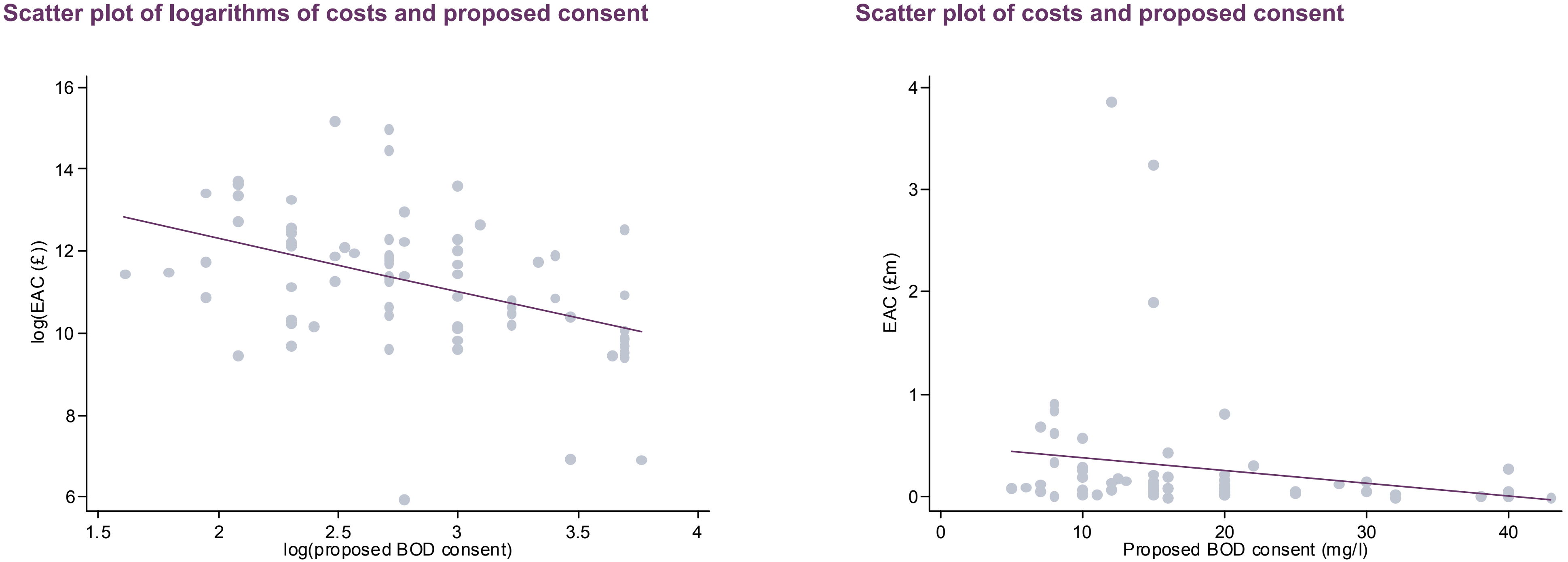


Scatter plot of logarithms of costs and proposed consent

Scatter plot of costs and proposed consent

Source: Oxera.

Figure A2.3 BOD: relationship between cost and proposed consent



Scatter plot of logarithms of costs and proposed consent

Scatter plot of costs and proposed consent

Source: Oxera.

A2.3 Frequency of treatment provision by company

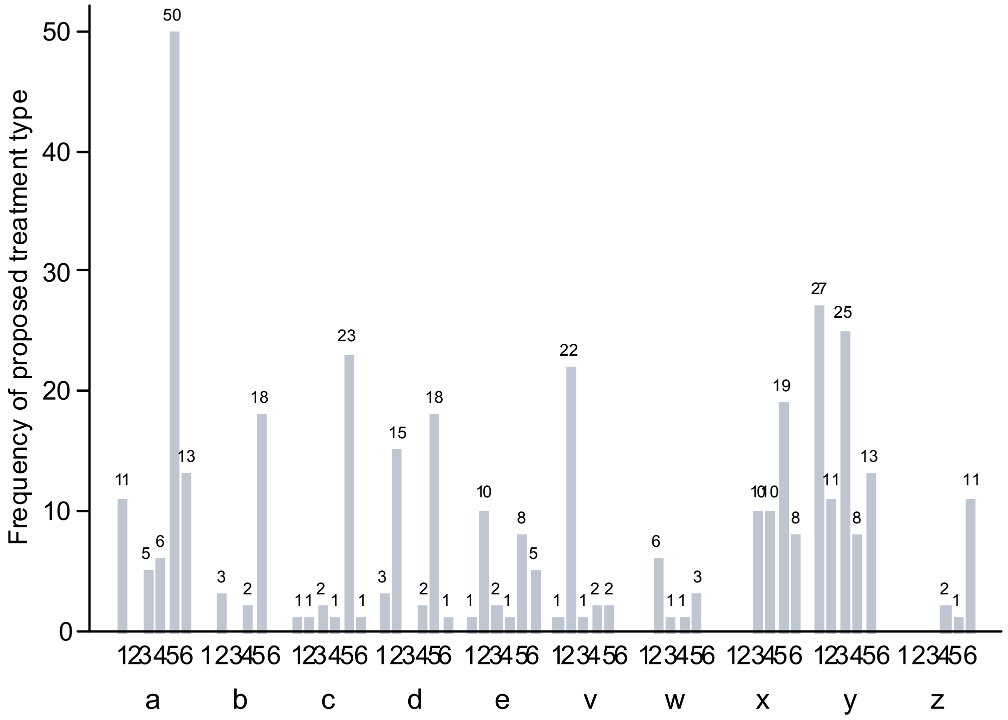
Figures A2.4 to A2.6 show the number of occurrences of each type of proposed treatment (categories 1 to 6) for each company (a to z). The bars display the frequency of each treatment.

The interpretation of the treatment type is as follows.

Table A2.10 Interpretation of treatment provision in Figures A2.4 to A2.6

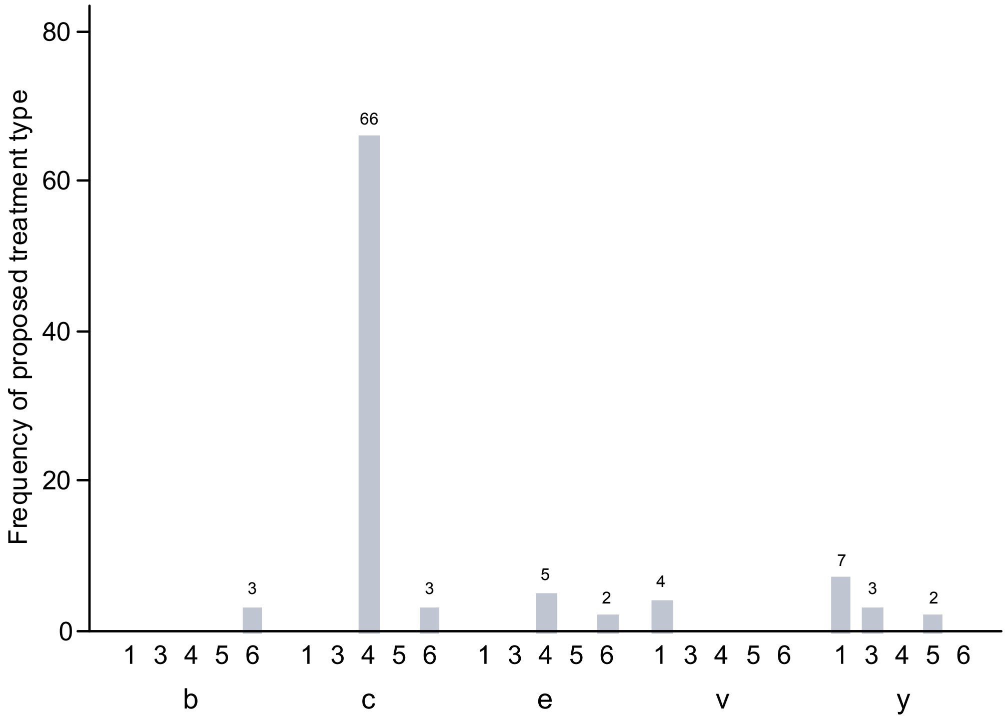
Code	Type of treatment provision
1	Secondary treatment activated sludge
2	Secondary treatment biological
3	Tertiary treatment A1
4	Tertiary treatment A2
5	Tertiary treatment B1
6	Tertiary treatment B2

Figure A2.4 Ammonia: frequency of proposed treatment provision by company



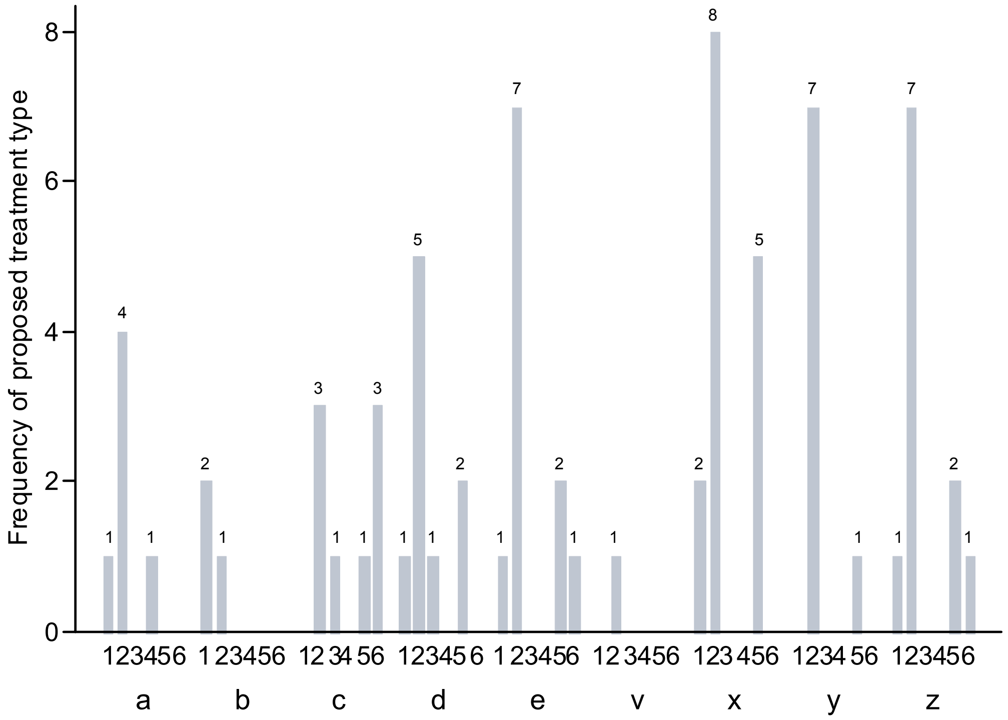
Source: Oxera.

Figure A2.5 Nitrates: frequency of proposed treatment provision by company



Source: Oxera.

Figure A2.6 BOD: frequency of proposed treatment provision by company



Source: Oxera.

A2.4 Diagnostic and outlier test results

Tables A2.11 to A2.13 summarise the results from the statistical testing of the cost functions for each parameter shown in Tables 4.1 to 4.3 respectively.

Table A2.11 Ammonia cost function: diagnostic tests and outlier analysis

Test	Test carried out	Critical value	Test statistic	Model correctly specified?
Functional form	RESET test	F(3, 349)	1.04	Yes
Heteroscedasticity	White test	Chi-Square (5)	6.6	Yes
Log versus linear	MacKinnon, White, and Davidson Pe test	F(1, 347)	2.43	Yes. Log linear specification is not rejected
Linear versus log	MacKinnon, White, and Davidson Pe test	F(1, 351)	145.56	Yes. Linear specification is rejected at any reasonable level of significance
Outlier analysis	Cook's Distance Leverage Welsch Distance			Yes, after removing outliers

Note: Although the residuals of the models do not pass the Jarque-Bera test for normality, a visual inspection of these suggests that these are approximately normally distributed.
Source: Oxera.

Table A2.12 Nitrates cost function: diagnostic tests and outlier analysis

Diagnostic test	Test carried out	Critical value	Test statistic	Model correctly specified?
Functional form	RESET test	F (3, 79)	1.39	Yes
Heteroscedasticity	White test	Chi-Square (2)	0.292	Yes
Log versus linear	MacKinnon, White, and Davidson Pe test	F (1, 78)	0.00	Yes. Log linear specification is not rejected
Linear versus log	MacKinnon, White, and Davidson Pe test	F (1, 79)	82.79	Yes
Outlier analysis	Cook's Distance Leverage Welsch Distance			Yes, after removing outliers

Note: Although the residuals of the models do not pass the Jarque-Bera test for normality, a visual inspection of these suggests that these are approximately normally distributed.
Source: Oxera.

Table A2.13 BOD cost function: diagnostic tests and outlier analysis

Test	Test carried out	Critical value	Test statistic	Model correctly specified?
Functional form	RESET test	F (3, 49)	0.33	Yes
Heteroscedasticity	White test	Chi-Square(9)	2.76	Yes
Log versus linear	MacKinnon, White, and Davidson Pe test	F (1, 52)	0.84	Yes. Log linear specification is not rejected
Linear versus log	MacKinnon, White, and Davidson Pe test	F (1, 58)	18.36	Yes. Linear specification is rejected at any reasonable level of significance
Outlier analysis	Cook's Distance Leverage Welsch Distance			Yes, after removing outliers

Note: Although the residuals of the models do not pass the Jarque-Bera test for normality, a visual inspection of these suggests that these are approximately normally distributed.
Source: Oxera.

Park Central
40/41 Park End Street
Oxford OX1 1JD
United Kingdom

Tel: +44 (0) 1865 253 000
Fax: +44 (0) 1865 251 172

Stephanie Square Centre
Avenue Louise 65, Box 11
1050 Brussels
Belgium

Tel: +32 (0) 2 535 7878
Fax: +32 (0) 2 535 7770

www.oxera.com