CEPA economics matters

PR24 Wholesale Base Cost Modelling

Ofwat

11 April 2023

FINAL REPORT

Important notice

Contents

EXECUTIVE SUMMARY

Over the course of recent price controls, Ofwat has developed a suite of econometric cost models that are used to assess water company base costs[1] submitted as part of the periodic review process. Ofwat's approach is well established but PR24 presents an opportunity to revisit and, where appropriate, refine the models.

CEPA was appointed to work alongside Ofwat to undertake such a review. This report identifies a selection of econometric cost models that performed well against our model selection criteria and that Ofwat can consider alongside or instead of the model specifications used to assess efficient wholesale base costs at PR19. This report will form an input into Ofwat's spring 2023 base cost modelling consultation.

Approach

- Our analysis has been undertaken using the dataset collected and published by Ofwat in November 2022 (v3) with some modifications.[2]
- We established a set of criteria to assess whether alternative model specifications merit consideration against the PR19 wholesale base cost models. The criteria test the engineering and economic logic, reliability, transparency and robustness of different model specifications (see Figure 2.1).
- To assess model robustness, we used a range of statistical tests drawing on those used to assess the models at PR19 (see Table 2.1 and Table 2.2).
- With two exceptions, we retained the levels of cost aggregation that were used at PR19 (see Table 4.1). In contrast to PR19, we created a new middle-up model for 'wastewater network plus' (combining sewage collection and sewage treatment costs) and we did not consider bioresources plus (sewage treatment + bioresources) models as Ofwat intends to assess bioresources and sewage treatment costs separately at PR24.
- We undertook the project in two phases. In the first phase we assessed a wide range of new or additional explanatory variables (see Table 4.2) as suggested by companies, Ofwat and our engineering advisers. Models that had a strong technical rationale and performed at least moderately well (amber or green rated) against a set of basic statistical criteria were progressed into a second stage of testing. In Phase 2, we further tested the robustness of the model specifications that passed our initial Phase 1 testing, using a range of statistical tests and sensitivity analysis.
- Our analytical start point was that the PR19 models form a good base for PR24 modelling. We tested PR19 models using the latest dataset and confirmed that the models continue to perform well in most cases. We therefore set a high bar for proposing that Ofwat consider a change as a result of Phase 2 testing.

[1] Base costs are routine year-on-year costs, which companies incur in the normal running of their businesses to provide a base level of service to customers and includes expenditure to maintain the long-term capability of assets, as well as expenditure to improve efficiency. Base costs typically consist of operating expenditure, maintenance capital expenditure and specific enhancement lines.

[2] For the dataset used in this analysis, Ofwat made some changes relative to the published v3 version including the addition of cost drivers used in this report and changes to the bioresources growth expenditure.

Phase 1 assessment

The following alternative or amended model specifications were progressed to Phase 2 modelling.

Table ES.1: Variables considered in our Phase 2 assessment for the water models (Water Resource Plus, Treated Water Distribution and Wholesale Water)

Model	WRP1	WRP2	TWD	WW1	WW2
Average Pumping Head (APH)[3]					
APH (TWD) used alone or in place of booster stations			✓	✓	✓
APH (TWD) used in addition to booster stations			✓	✓	✓
APH (all) used alone or in place of booster stations				✓	✓
APH (all) used in addition to booster stations				✓	✓
Alternative density drivers					
MSOA[4]	✓	✓	✓	✓	✓
LAD from MSOA	✓	✓	✓	✓	✓
Properties per length of mains	✓	✓	✓	✓	✓
Alternative water treatment complexity drivers					
Alternative treatment complexity driver (SW and GW 3-6 bands	✓			✓	
Alternative weights for weighted average complexity		✓			✓
Alternative scale drivers in wholesale water models					
Length of mains				✓	✓
Water resources drivers					
% Distribution Input (DI) from reservoirs	✓	✓		✓	✓
% DI from pumped reservoirs	✓	✓		✓	✓
Time trend					
Time trend			✓		

[3] We have used APH TWD (treated water distribution) for the TWD and WW models, and APH (all, i.e., resources, transport, treatment and treated water distribution) for the WW models.

[4] Middle layer Super Output Area (MSOA) estimates are sourced from the Office of National Statistics (ONS) and provide a more granular view of density compared to the Local Authority Distract area data underpinning the PR19 weighted average density measure.

Table ES.2: Variables considered in our Phase 2 assessment for the wastewater models

Model	SWC1	SWC2	SWT1	SWT2	WWWN+
Rainfall					
Urban rainfall per sewer length (log)	✓	✓			✓
Treatment complexity					
BOD consents ≤ 7mg/l (%)			✓	✓	
Network topography and coastal population					
'Coastal' population (%)			✓	✓	
Pumping capacity per sewer length			✓	✓	
Sewage treatment economies of scale					
Sewage treatment work size cut-offs				✓	✓
Weighted Average Band (Anglian Water)				✓	✓
Weighted Average Band (CEPA)				✓	✓
Weighted average treatment size				✓	✓
Density					
MSOA density, aggregated at LAD level		✓			
MSOA density		✓			

Table ES.3: Variables considered in our Phase 2 assessment for the bioresources models

Model	BR1	BR2
Alternative density drivers		
MSOA (population)		
LAD from MSOA (population)	✓	
Properties per length of mains	✓	
Economies of scale in sludge treatment		
CEPA size bands	✓	✓
Weighted average bands (WAB)	✓	✓

Phase 2 assessment

In Phase 2, we focused our analysis in three areas:

1) Statistical performance against the following tests

- The **explanatory power (R^2)** and **parameter significance (t-test)**.
- Other statistical tests such as **RESET**, **VIF**, **normality** and **pooling** tests.
- **Sensitivity analysis** by excluding years or companies from the sample.

2) We report the **range and stability of efficiency scores** relative to the PR19 model specifications. While a company's efficiency score, relative to the PR19 model specifications, is not an indication that an alternative model performs better or worse than the PR19 specifications, it provides an indication of the impacts of selecting alternative models. If the overall range of efficiency scores is larger than the comparative PR19 model specification, this could indicate the presence of issues in the underlying model.

3 As noted in our approach above, in Phase 2 we compare the performance of alternative models to the PR19 model specifications and, apply a high bar for recommending alternative options as potential additions or replacements to the PR19 model suite. In order to consider recommending a change to Ofwat, we would generally expect a new model to:

- pass the high and medium importance statistical tests that we set out in Table 2.1;
- prove robust under the range of sensitivity tests; and
- meet the wider selection criteria established in Figure 2.1 including technical rationale, data quality and transparency.

Recommendations

Our Phase 2 testing suggests that only a moderate level of change to the PR19 models should be considered. Our recommendations to Ofwat are set out below:

For the PR24 base cost **water modelling**, we recommend that Ofwat should consider the following models:

- The PR19 treated water distribution models and wholesale water models with the inclusion of **APH** if the APH data is sufficiently robust once an additional year of data becomes available. Ofwat should also consider whether there is an engineering rationale for including boosters per length alongside APH.
- An alternative wholesale water model with **length of mains** used as the scale driver, though the impact on the boosters per length variable of including length of mains and possibly APH in the WW models should be considered further. An **alternative density driver** (MSOA, LAD aggregated from MSOA or properties per length of mains) for all water models. Ofwat should consider whether this would improve data quality and increase the robustness of the density variable. These alternative density drivers can avoid some of the issues encountered with the PR19 density measure such as relying on mapping of company areas to LADs and sensitivity to changes in ONS boundaries. This can improve the transparency and reliability of the dataset.

For the PR24 base cost **wastewater modelling**, we recommend that Ofwat should consider the following models:

- The PR19 sewage collection models with the addition of **normalised urban rainfall.** Similar to the water models, Ofwat should also consider whether an alternative density driver (MSOA or LAD aggregated from MSOA) would improve the data quality and robustness of the density variable.
- The PR19 sewage treatment models with **alternative economies of scale** drivers. Ofwat should consider which alternative driver is best supported by technical rationale. Our assessment indicates that the drivers that perform the best against our model selection criteria are:
    - For economies of scale drivers based on threshold size (SWT models only):
        - load treated in STWs ≥ 100,000.
    - For weighted average economies of scale drivers (SWT and WWWN+ models):
        - CEPA's weighted average band; and
        - Weighted Average Treatment Size
- Wastewater network plus models that include:
    - load;
    - pumping capacity per sewer length;
    - load treated with ammonia consent ≤ 3mg/l;
    - urban rainfall per sewer length; and
    - a weighted average band or weighted average treatment size variable.

For the PR24 **base cost bioresources models**, we recommend that Ofwat should consider models that include:

- a scale driver (sludge produced);
- an economies of scale driver: load treated in size bands 1-3, sewage treatment works per number of properties or the CEPA WAB variable; and
- potentially a density driver (with LAD from MSOA being the best performing density variable in the BR models based on our analysis) if there is a sufficiently strong technical rationale to suggest density should be used either in addition or instead of an economies of scale driver.

1. INTRODUCTION

Over the course of recent price controls, Ofwat has developed a suite of econometric cost models that are used to assess water company base costs[5] submitted as part of the periodic review process. Ofwat's approach is well established but PR24 presents an opportunity to revisit and, where appropriate, refine the models. CEPA was appointed to work alongside Ofwat to undertake such a review. This report presents the results and will form an input into Ofwat's spring 2023 base cost modelling consultation.

1.1. Scope of work

The purpose of CEPA's work has been to assist Ofwat in developing a suite of wholesale base cost models that can be used to determine efficient wholesale base cost allowances at PR24.

We reviewed the base cost model specifications used at PR19 and considered potential improvements, where appropriate. As part of this, we carried out a high-level review of the cost assessment dataset published by Ofwat, and selected a set of econometric models to be evaluated as part of this review. The list of models considered in this study accounted for suggestions put forward to Ofwat by water companies, as well as Ofwat's own suggestions and input from our engineering panel. We also established a set of criteria to assess whether the model specifications tested are suitably robust and valid for the purposes of informing or setting cost baselines as part of the price review. We then ran and independently assessed the performance of the models included within the scoping exercise against the assessment criteria.

We undertook modelling in two phases, with models that passed the initial selection in Phase 1 being progressed to further testing in Phase 2. CEPA was supported in the review by a panel of engineers and econometricians who provided specialist input into the process.

In line with Ofwat's wider approach to PR24 cost assessment, we take the PR19 wholesale base cost models as the starting point for our analysis and in considering alternative specifications, we have had regard to how new models perform against the PR19 suite. Issues related to the definition of base costs, pre-modelling adjustments to costs and the level of cost aggregation modelled were not part of the scope of this engagement, except in the following cases:

- We considered 'wastewater network plus' models (combining sewage collection and sewage treatment costs) which were not part of the PR19 suite of models.
- We did not consider 'bioresources plus' models (combining sewage treatment and bioresources costs), as Ofwat proposes to assess bioresources costs separately from other wholesale wastewater costs at PR24.

This report identifies a selection of econometric cost models that performed well against our model selection criteria. Ofwat can consider these models alongside or instead of the model specifications used to assess efficient wholesale base costs at PR19.

1.2. Structure of report

The remainder of the report is structured as follows:

- Section 2 describes our approach to the development and assessment of models, including the data and the criteria used to assess the robustness of the models.

[5] Base costs are routine year-on-year costs, which companies incur in the normal running of their businesses to provide a base level of service to customers and includes expenditure to maintain the long-term capability of assets, as well as expenditure to improve efficiency. Base costs typically consist of operating expenditure, maintenance capital expenditure and specific enhancement lines.

- Section 3 describes the wholesale base cost models used by Ofwat at PR19 and presents statistical results based on running the PR19 model specifications with the latest available data.
- Section 4 sets out the alternative model specifications that we have analysed in this study, including the levels of cost aggregation and explanatory variables used.
- Section 5 presents the findings of our initial assessment of the models in the first phase of the project.
- Section 6 presents the results of the detailed assessment undertaken in the second phase of the project.
- Section 7 discusses a set of options for Ofwat's consideration at PR24 based on the model specifications that performed best against our selection criteria.

Detailed model results are presented in the annexes.

2. METHODOLOGY

In this section we present our approach to developing and assessing econometric models. As with all our work, this builds on the methods established for PR19.

To maintain transparency and objectivity, we undertook a series of analytical steps prior to starting the modelling process. These included:

- Reviewing the data inputs required for the models.
- Identifying the model selection criteria used to assess models, including statistical tests.
- Setting out the level of cost aggregation (i.e., the dependent variables) and the cost drivers/explanatory variables to be included in the models.
- Establishing, with engineering input, the rationale for including a model specification within our scope.

We discuss the first two points in more detail in the remainder of this section and the last two steps in Section 0.

2.1. Data

We have used data collected and published by Ofwat in autumn 2022. The dataset includes outturn data up to the end of 2021-22 submitted by companies as part of the Annual Performance Report (APR) process as well as additional information collected by Ofwat from companies to support its approach to base cost assessment at PR24.[6] Our analysis has been undertaken using the dataset published by Ofwat in November 2022 (v3).

The data used was consolidated and validated by Ofwat. Prior to conducting our modelling exercise, we undertook a high-level review to identify general trends and potential anomalies in the data, but we did not perform an independent in-depth quality assurance. Where we identified potential data issues, we raised them with Ofwat for further investigation. In a small number of cases this led to a change in the data.

2.2. Model selection criteria

We established a set of criteria to assess whether alternative model specifications merit consideration against the PR19 wholesale base cost models. A clear set of evaluation criteria helps to objectively demonstrate whether model results are suitably robust and valid for the purposes of informing or setting cost baselines as part of the price review. A set of assessment criteria should test the logic, reliability, transparency and robustness of different model specifications.

We use the following criteria to assess model performance. This is in line with Ofwat's principles for PR24 base cost assessment, and the model selection criteria CEPA previously used when advising Ofwat at PR19 and Ofgem at RIIO-GD2 and RIIO-ED2.[7]

---

[6] The PR24 cost assessment datasets published by Ofwat can be found at: https://www.ofwat.gov.uk/regulated-companies/price-review/2024-price-review/pr24-cost-assessment-datasets/

[7] Ofwat model criteria (p.7-8): https://www.ofwat.gov.uk/wp-content/uploads/2022/12/PR24_final_methodology_Appendix_9_Setting_Expenditure_Allowances.pdf

CEPA model criteria PR19 (p.124): https://www.ofwat.gov.uk/wp-content/uploads/2018/03/CEPA-cost-assessment-report.pdf

Figure 2.1: Model assessment criteria

Economic/technical rationale	• Do the model specifications and results have a clear economic/technical rationale? • Are the selected explanatory variables consistent with an engineering view? • Are the stated coefficients consistent with a priori expectations of magnitude and signs of estimated coefficients?
Robustness	• Does the model pass the statistical tests (see 2.4.1 below)? • Is the model sensitive to the underlying assumptions? We test this through sensitivity analysis such as removing one year of data or removing companies.
Transparency	• Are the data used and results transparent and easy to interpret? • Is the model understandable and intuitive? This should consider an appropriate balance between simplicity and complexity if complexity brings a significant improvement in the performance of the model.
Data requirements	• Is the data used in the model available to all stakeholders? Ofwat data has been shared with and validated by the companies. • How reliable is the available data used in the model?
Consistency with regulatory objectives and policy	• Is the model consistent with and does it create incentives that align with regulatory objectives? • Does the model create perverse incentives or distort companies' behaviour? • Does the model rely on exogenous cost drivers that are outside of company control?

In this report our main focus is on three criteria – economic/technical rationale, statistical robustness and consistency with Ofwat's regulatory objectives. The data requirements criterion has two components – availability and reliability. We consider the availability component to be met by the shared dataset that Ofwat has developed. We consider the data quality component on a case-by-case basis where there were reasons for concern regarding data quality for specific items in Ofwat's dataset, or when assessing models that utilise new data sources or a particular variable suggested by a company.

2.3. Statistical and robustness test

To assess model robustness, we used a range of statistical tests drawing on those used to assess the models at PR19. Ideally, the final models selected would pass all model evaluation criteria and tests they are submitted to. However, setting such a high standard would make it very difficult to develop models at all. We therefore set out below our view of the relative importance of each of the core tests considered in our assessment.

We grouped the statistical and robustness tests by importance as follows:

- **High:** Tests and criteria that when failed would raise serious concerns about using a model.

- **Medium:** Tests and criteria that, when failed, would raise some concerns about using the model but the model could be used with caution if it passes other tests.
- **Low:** Tests and criteria that, when failed, would raise relatively limited concerns about using the model

Table 2.1: Statistical tests

Importance	Test	Description
High	*Goodness of fit (overall R^2)*	*If a model fails to explain a substantial share of the variation in costs of the industry, it would be inappropriate to use it for the estimation of the costs going forward. This test provides the R-squared for Random Effects (i.e., (STATA's R-squared overall, which is calculated as correlations squared).*
	Statistical significance of individual parameters (t-test)	*If one or more of the coefficients in the model fails this test, we cannot rule out that the relationship being identified between the cost driver and costs under consideration is spurious (i.e. the coefficient could be zero).* *Parameters could fail this test because there is no relationship between the cost driver and the costs but also due to limitations in the data or multicollinearity.*
Medium	*RESET test*	*Tests whether there are non-linearities in the data that have not been captured adequately by the estimated model. Failing this test may indicate that the data could be better fitted using a different shape (e.g. quadratic). However, this is not to say that a linear assumption is automatically wrong but that other options should be explored. If alternative specifications using non-linear terms in the model do not yield satisfactory results, then the failure of the RESET test on its own may not be a valid justification to dismiss a model. This is particularly the case if the model offers useful information from an economic or engineering perspective.*
	Pooling test	*To use a panel data estimation method, we need to assume that the coefficients being estimated are stable over time. If this assumption fails, panel data analysis may not be appropriate. This can be tested with the Pooling test.* *The Pooling test examines whether the coefficients on each individual year of data in the model are significantly different from the coefficients of the model running from 2012-2022 (whole sample).*
	Variance Inflation Factor (VIF)	*Models with highly correlated variables have a higher risk of introducing the same information into the model. Multicollinearity can cause some relevant variables to be insignificant or with unexpected values. Models with a max and/or mean VIF above 10 are considered to have a relatively high risk of suffering from multicollinearity. Even when multicollinearity exists, the models can still be considered as valid if the different variables are significant, and the overall result appears to be consistent with any initial expectations for the relevant effect.*
Low	*Heteroskedasticity*	*If a model fails the heteroskedasticity test, it means that the variance of the errors is not equal for all observations. It typically occurs when the variation in the residuals is very different over time.* *We assign low level of importance to this test, as we use clustered robust standard errors to control for potential heteroskedasticity. Therefore, the model can still be used if it fails the test, as failure does not affect the robustness of the model.*
	Normality	*The test for normality is used to assess whether the residuals are normally distributed. Failure of this test affects statistical inference. However, this does not introduce a bias in the estimated coefficients. We therefore apply a low level of importance to this test.* *We only conduct this test for OLS.*
	Breusch-Pagan LM test	*Test of pooled OLS versus RE. If the models fail this test, the effects are like the ones discussed above for heteroskedasticity i.e., the results are still robust, but they do not achieve all the positive properties that are normally associated with an OLS estimate. Failure of the test would*

Importance	Test	Description
		provide an indication that random effects is preferred over Pooled OLS estimation.

We also consider the stability of the models and whether there are specific observations (or specific companies) that drive the overall result. We apply a range of robustness tests and metrics to test the sensitivity and stability of results. These are discussed in Table 2.2 below.

Table 2.2: Robustness tests

Test	Description
Excluding most/least efficient or outlier company	*We apply a RAG assessment based on changes in sign and significance of coefficients:* • *Red (R): the estimated coefficients present changes in both significance and sign;* • *Amber (A): the estimated coefficients present some changes in significance but not in sign; and* • *Green (G): the estimated coefficients do not present changes in significance or sign.*
Removing individual years of the sample/truncating the sample period	*RAG assessment based on changes in sign and significance of coefficients, as above.*
Range of efficiency scores	*We report the range of the minimum and maximum efficiency scores from each model.* *A large range of efficiency scores could indicate the presence of issues in the underlying model.*
Stability of efficiency scores and rankings relative to PR19 model specifications	*We check if the efficiency score of any company is 5 percentage points (pps) lower or higher than the updated PR19 model. We also check if the most/least efficient company in the updated PR19 model are in the three most/least efficient companies in the alternative model.* *Large changes in efficiency scores relative to the PR19 model specifications are not an indication that an alternative model performs better or worse than the PR19 specifications. However, it provides an indication of the magnitude of the impact on individual companies to be expected if an alternative model is adopted.*

2.4. Quality Assurance (QA)

QA has been undertaken on an ongoing basis throughout the project; all modelling results were reviewed, and sense checked by relevant team members (including the project manager and an econometrician) to maintain our high quality standards. We also involved our engineering adviser and Ofwat's Academic Econometrics Adviser in the analysis as it was progressed.

For the draft and final report we undertook the following:

- **Internal CEPA QA.** A CEPA staff member outside of the immediate project team conducted an end-to-end QA process checking and rerunning the STATA-do file created by the CEPA project team in order to validate the analysis from STATA to Excel to Final Report. The findings of this review were shared with the project team and any required corrections were made by them prior to the report being signed off and issue to Ofwat

- **External QA.** *Ofwat's Academic Econometric Advisor*, has had access to our working analysis and results files throughout and has advised on technical matters including the Phase 2 sensitivity tests included in our STATA-do file. The high level assurance statement provided by Ofwat's Academic Econometric Advisor extends to his review of our work.

3. OFWAT'S PR19 MODEL SPECIFICATIONS AND USE AT PR24

The first stage of our analysis was to re-run the PR19 models including more recent data. This section describes that process.

3.1. Ofwat's approach at PR19

In December 2019 Ofwat published its PR19 final determinations, including its approach to wholesale water and wholesale wastewater base cost assessment.[8] Ofwat's view of efficient base costs drew heavily on econometric cost modelling using companies' actual expenditure over 2011-12 to 2018-19. Ofwat used the following levels of aggregation:

- **Wholesale water**: Ofwat used econometric models to benchmark costs at water resources plus (WRP, 2 models), treated water distribution (TWD, 1 model) and wholesale water (WW, 2 models) levels. "Water resources plus" consist of water resources, raw water distribution and water treatment combined. Wholesale water consists of water resources plus and treated water distribution.
- **Wholesale wastewater**: Ofwat used models to benchmark costs at sewage collection (SWC, 2 models), sewage treatment (SWT, 2 models), bioresources (BR, 2 models) and bioresources plus (BRP, 2 models) levels. "Bioresources plus" consists of bioresources and sewage treatment combined.

Ofwat tested models using two methods, ordinary least squares (OLS) and random effects. The final determination models were based on random effects. Ofwat concluded that these models performed better statistically than under the OLS method and the Breusch Pagan test consistently supported using random effects over the OLS. At PR19 Ofwat's starting point was the Cobb-Douglas (or "constant elasticity") functional form.[9] This model assumes that scale or density effects are constant. That is, a percentage change in the explanatory variable (for example scale or density) results in the same percentage change in costs for all companies. Ofwat's PR19 wholesale water models included a non-linear density term, to capture the U-shaped relationship between base costs and density.[10]

In summary the models indicated that the key factors driving costs in wholesale water and wastewater are:

- scale/output/volume – larger scale/output/volume drives higher total costs;
- density of the network and/or the population served;
- topography of a company's geographical area – which drives factors such as pumping and therefore energy use;
- treatment complexity – higher complexity drives higher costs; and
- size of treatment works – larger size drives lower unit cost due to economies of scale at the treatment level *(for wastewater models only).*

Ofwat's PR19 models performed well, and withstood the independent scrutiny of the CMA, with one minor adjustment. The CMA included a squared density variable in one of the sewage collection models (SWC2), similar

[8] https://www.ofwat.gov.uk/wp-content/uploads/2019/12/PR19-final-determinations-Securing-cost-efficiency-technical-appendix.pdf

[9] Cobb-Douglas is a standard function form used in cost assessment literature that places weights on the input factors (i.e. cost drivers). When in a log-linear form, Cobb-Douglas models assume an additive approach across the different cost drivers and it allows marginal costs to vary and for coefficients to be interpreted as cost elasticities. Cobb-Douglas models are relatively easy to replicate and interpret.

[10] At lower levels of density, scale economies are strong and therefore increasing density reduces costs. However, the positive effect of the quadratic term suggests that as density rises its negative impact on costs decreases, ultimately becoming positive at high values of density.

to Ofwat's approach to capturing density in the PR19 wholesale water models.[11] Our analytical start point therefore was that the PR19 models form a good base for PR24 modelling and we set a relatively high bar for proposing that Ofwat consider a change.

Ofwat recognises that there are practical limitations on the use of statistical modelling in cost assessment. At PR19, Ofwat set efficient base cost allowances by triangulating cost estimates across a set of models to mitigate the risk of error and bias from over reliance on any particular model. [12] We test a range of models to assess the case for revising PR19 models, but we expect that Ofwat will similarly triangulate across models in its PR24 decision making process.

3.2. Scope of base costs at PR24

Ofwat asked us to develop econometric cost models based on the following definitions of wholesale base costs:

- **wholesale water**: opex plus capital maintenance plus addressing low pressure plus network reinforcement expenditure;
- **wholesale wastewater network plus**: opex plus capital maintenance plus transferred private sewers and pumping stations plus network reinforcement plus reducing risk of sewer flooding enhancement expenditure plus enhancement opex for cost categories where Ofwat were relatively certain that the costs are ongoing; and
- **bioresources**: opex plus historical quality enhancement opex plus capital maintenance plus sludge growth enhancement expenditure.

Ofwat also made the following pre-modelling data adjustments to facilitate accurate comparison between companies and over time.

Table 3.1: Pre-modelling adjustments for base costs at PR24

Adjustment	Description
Unmodelled costs	Ofwat excluded costs that will be treated as unmodelled base costs at PR24. Unmodelled base costs include pension deficit recovery costs, business rates, abstraction and discharge charges (water only), costs associated with the Traffic Management Act, statutory water softening costs, wastewater Industrial Emissions Directive (IED) operating costs, and third-party costs.
Atypical expenditure adjustment	Ofwat has decided to include atypical expenditure in modelled base costs by default at PR24. But atypical costs that relate to fines/penalties, accounting adjustments, costs associated with referrals to the Competition and Markets Authority (CMA), and truly one-off atypical costs that are unlikely to be repeated (e.g. costs incurred in preparation for the introduction of retail competition for business customers) will be excluded.
Principal use adjustment	Principal Use of Assets (PUA) accounting treatment was introduced in 2015-16. This means that the base costs of assets used in more than one price control are allocated to the largest of the relevant price controls. Compensating accounting transactions are then made by the other price controls to recompense the price control of principal use. Ofwat found that companies had not always made the correct principal use adjustments to recompense the price control of principal use. An adjustment has been applied to base costs from 2015-16 onwards to correct for this issue.

[11] CMA (17 March 2021), Price Determination, p.161.
https://assets.publishing.service.gov.uk/media/60702370e90e076f5589bb8f/Final_Report_---_web_version_-_CMA.pdf

[12] Ofwat (December 2019), PR19 final determinations, securing cost efficiency technical appendix, p.36.

Adjustment	Description
Bioresources and sewage treatment 'backcasting' adjustment	This adjustment aims to account for Ofwat's updated guidance on how to allocate the costs of sludge liquor treatment, energy generation and overheads between bioresources and sewage treatment.
Developer services base cost adjustment	Ofwat is excluding site-specific developer services expenditure from the scope of modelled base costs at PR24. This adjustment ensures that historical developer services costs that had previously been reported in operating and capital maintenance expenditure are not included in modelled base costs. The adjustment was informed by data submitted by water companies to Ofwat in summer 2022.

3.3. Rerunning PR19 model specifications with latest dataset

To establish a baseline for our assessment of alternative model specifications and to confirm that the PR19 model specifications remain fit for purpose, we ran Ofwat's PR19 model specifications (as modified by the CMA) using the latest dataset and the updated definition of base costs established for PR24, as set out in Section 3.2. We assessed these updated models against our model selection criteria. We also assessed the following:

- Are there changes in statistical robustness compared to PR19?
- What has changed in the underlying costs/driver data that could have caused this change?
- Is there a need to amend the models to control for these changes?

We found that the PR19 WRP1, TWD1, WW1 and WW2 models model specifications re-run with the updated dataset and cost definitions show similar results in terms of model robustness compared to PR19. All coefficients on the independent variables remain statistically significant on a 5% basis, the signs are as expected and the R-squared does not change materially.

However, the WRP2 model performs less well than at PR19, as the estimated coefficient on weighted average treatment complexity is no longer significant on a 10% basis, while it remains so in the WW2 model. As a result, we conducted further sensitivity testing (which is discussed in Section 6.1) to understand the robustness of these results and considered alternative weighted average treatment complexity variables as part of our assessment of alternative model specifications.

Table 3.2: Statistical results for PR19 wholesale water model specifications (PR19 results vs. results with data series extended to end of 2021/22)

Model	WRP1		WRP2		TWD		WW1		WW2	
Version	**PR19**	**PR24**	**PR19**	**PR24**	**PR19**	**PR24**	**PR19**	**PR24**	**PR19**	**PR24**
Properties (log)	1.033*** {0.000}	1.074*** {0.000}	1.030*** {0.000}	1.069*** {0.000}			1.036*** {0.000}	1.071*** {0.000}	1.024*** {0.000}	1.059*** {0.000}
Water treated in bands 3-6 (%)	0.008*** {0.000}	0.006*** {0.000}					0.006*** {0.000}	0.004*** {0.000}		
WAD - LAD (log)	-1.451*** {0.001}	-1.614*** {0.000}	-0.958*** {0.024}	-1.412*** {0.005}	-3.338*** {0.000}	-2.946*** {0.000}	-2.371*** {0.000}	-2.094*** {0.000}	-1.939*** {0.000}	-1.832*** {0.000}
Squared WAD - LAD (log)	0.091*** {0.004}	0.101*** {0.000}	0.055* {0.064}	0.087*** {0.009}	0.266*** {0.000}	0.235*** {0.000}	0.168*** {0.000}	0.147*** {0.000}	0.137*** {0.000}	0.128*** {0.000}
Weighted average complexity (log)			0.444*** {0.005}	**0.377** **{0.123}**					0.533*** {0.000}	0.430*** {0.001}
Lengths of main (log)					1.055*** {0.000}	1.077*** {0.000}				
Boosters per length (log)					0.570*** {0.000}	0.437*** {0.002}	0.361*** {0.004}	0.335** {0.032}	0.324*** {0.001}	0.334** {0.019}
Constant	-5.307*** {0.001}	-5.093*** {0.000}	-6.979*** {0.000}	-5.805*** {0.000}	6.782*** {0.000}	4.723*** {0.002}	-0.331 {0.801}	-1.565* {0.074}	-1.948* {0.053}	-2.589*** {0.001}
R-squared	0.93	0.917	0.92	0.907	0.96	0.957	0.97	0.97	0.98	0.971

Note: this table uses PR19 CMA final determination results.

A similar picture emerges for the PR19 wholesale wastewater model specifications. We found similar results in terms of model performance, sign and magnitude of coefficients, with only marginal differences compared to PR19:

- The R-squared decreased marginally across most models.
- The coefficient in SWC1 and SWC2 are statistically significant on a 5% basis. The coefficient on load treated in size band 6 (%) in SWT2 is slightly less significant, but remains significant on a 10% basis.
- The coefficient on load treated in bands 1-3 is no longer significant in the SWT1 model. We considered alternative economies of scale drivers in sewage treatment works (STW) as part of our assessment of alternative model specifications.

Table 3.3: Statistical results for PR19 wholesale wastewater model specifications (PR19 results vs. results with data series extended to end of 2021/22)

Model	SWC1		SWC2		SWT1		SWT2	
Version	**PR19**	**PR24**	**PR19**	**PR24**	**PR19**	**PR24**	**PR19**	**PR24**
Sewer length (log)	0.839*** {0.000}	0.804*** {0.000}	0.830*** {0.000}	0.859*** {0.000}				
Load (log)					0.779*** {0.000}	0.653*** {0.000}	0.781*** {0.000}	0.658*** {0.000}
Load treated in size bands 1-3 (%)					0.042** {0.016}	**0.029** **{0.211}**		
Load treated in size bands 6 (%)							-0.012** {0.025}	-0.009* {0.097}
Pumping capacity per length (log)	0.291* {0.088}	0.344** {0.012}	0.501** {0.012}	0.604*** {0.000}				
Load with ammonia consent < 3mg/l (%)					0.004*** {0.000}	0.006*** {0.000}	0.004*** {0.000}	0.006*** {0.000}
Properties per length (log)	0.986*** {0.009}	1.043*** {0.000}						
Weighted average density (log)			-2.683** {0.040}	-2.480** {0.021}				
Square weighted average density (log)			0.194** {0.024}	0.181*** {0.010}				
Constant	-8.030*** {0.000}	-7.956*** {0.000}	4.845 {0.296}	3.606 {0.395}	-5.211*** {0.000}	-3.734*** {0.004}	-4.118*** {0.000}	-2.965*** {0.000}
R-squared	0.934	0.917	0.913	0.895	0.873	0.854	0.864	0.855

Note: this table uses PR19 CMA final determination result

The bioresources models, rerun using the updated dataset and cost definitions, perform less well than at PR19. In BR1, the weighted average density coefficient is no longer statistically significant. In BR2, the STWs per property coefficient is likewise no longer statistically significant. Nonetheless, the R-squared does not change significantly. We consider alternative density and economies of scale drivers alongside other model specifications as part of our assessment.

Table 3.4: Statistical results for PR19 bioresources model specifications (PR19 results vs results with updated data series and cost definition)

Model	BR1		BR2	
Version	**PR19**	**PR24**	**PR19**	**PR24**
Sludge produced (log)	1.294*** {0.000}	1.172*** {0.000}	1.313*** {0.000}	1.134*** {0.000}
WAD - LAD (log)	-0.348** {0.016}	-0.133 {0.267}		
Load treated in STWs bands 1-3 (%)	0.054** {0.027}	0.063** {0.011}		
STWs per property (log)			0.447* {0.052}	0.275 {0.174}
Constant	-0.081 {0.921}	-0.912 {0.310}	1.182** {0.046}	0.808 {0.316}
R-squared	0.818	0.82	0.788	0.784

Note: this table uses PR19 CMA final determination results

4. ALTERNATIVE MODEL SPECIFICATIONS

This section presents the different levels of cost aggregation that we use as the dependent variable in our modelling (taking the base cost definitions set out in Section 3.2.), and then lists the explanatory variables we evaluate in our Phase 1 assessment.

4.1. LEVEL OF COST AGGREGATION

At PR19, Ofwat used models at differing levels of cost aggregation including **top-down** aggregated models for wholesale water, **middle-up** models for 'water resources plus' (combining water resources, raw water distribution and water treatment costs) and 'bioresources plus' (combining sewage treatment and bioresources costs), and **bottom-up** disaggregated models for treated water distribution, sewage collection, sewage treatment and bioresources.

The use of models at different levels of cost aggregation accounts for trade-offs between disaggregated and more aggregated cost models. The disaggregated models allow a wider range of cost drivers to be captured in the modelling whereas more aggregated models capture interactions between different services and mitigate potential cost allocation issues.

We were asked by Ofwat to consider models at the same level of cost aggregation as for PR19 but with two exceptions:

- We created a new middle-up model for 'wastewater network plus' (combining sewage collection and sewage treatment costs) which was not part of the PR19 suite of models.
- We did not consider bioresources plus (sewage treatment + bioresources) models as Ofwat intends to assess bioresources and sewage treatment costs separately at PR24.

Table 4.1 below summarises the levels of cost aggregation considered in our analysis.

Table 4.1: Levels of cost aggregation modelled split by top down, middle-up and bottom-up models

Top-down	Middle-up	Bottom-up
Wholesale water	*Water resources plus (water resources + raw water distribution + water treatment)*	*Treated water distribution*
	Wastewater network plus (sewage collection + sewage treatment)	*Sewage collection*
		Sewage treatment
		Bioresources

4.2. POTENTIAL EXPLANATORY VARIABLES

We considered a number of variables that were suggested by companies in response to Ofwat's base cost assessment consultation and by Ofwat's and CEPA's engineering advisers. We present the cost drivers and explanatory variables that we model, and the rationale for doing so, in the table below.[13]

[13] The difference between cost drivers and explanatory variables is that a cost driver represents the underlying factors driving company costs, for example, the concentration of contaminants in a water source. As actual cost drivers may be difficult to measure, an explanatory variable is generally a variable for which data is available and that constitutes a (partial) proxy for the cost driver that it tries to capture. In this example, an explanatory variable for the concentration of contaminants may be percentage of water coming from different sources. Given that different sources are expected to have different quantities of contaminant load, these variables could be used as a proxy for the overall cost driver.

Table 4.2: Alternative explanatory variables tested

Cost driver	Explanatory variable	Explanation	Models included
Wholesale water models			
Network topology	• *Average pumping head (APH)*	*The APH variable provides a proxy for the energy requirements that companies face. The PR19 models used booster pumping stations per length of mains instead of APH.*[14] *While the former controls for topography (i.e., hilly areas require more boosters) and asset intensity, APH is also used to capture the volume of water (i.e., driven by topography and population size) and water pressure.* *The expectation is that the greater the level of pumping required, the greater the amount of energy used and, as a result, the higher the total cost incurred by the company.* *Most pumping costs are related to treated water distribution so we would expect APH to be most relevant for explaining TWD costs.*	• *WRP* • *TWD* • *WW*
Alternative density variables	• *MSOA*[15] • *Local Authority District (LAD) from MSOA* • *Properties per length of mains*	*Density can have two counteracting effects on costs: on one hand, dense areas may have high costs due to higher labour, property and access costs; on the other hand, dense areas can be served by larger treatment works which incur lower unit costs.* *The alternative density variable (MSOA) uses more granular ONS data than was available at PR19, avoiding issues created by the merging of LADs over the time series covered by the base models. We also tested a density variable that uses MSOA data to create LAD level density (referred to as "LAD from MSOA") and a simple properties per length of mains variable, to mirror the variable used in PR19 SWC1 model (properties per length of sewers).* *In addition, we also tested removal of the squared density term from the PR19 model, as suggested by Severn Trent Water.*[16]	• *WRP* • *TWD* • *WW*
Water treatment complexity	• *Alternative treatment complexity bands*	*Water treatment complexity can drive higher costs through the resources required e.g., power and chemicals. The complexity of treatment is a function of both the quality of the raw water source and treated water quality requirements.*	• *WRP* • *WW*

[14] The CMA confirmed that, when judging on statistical significance, booster pumping station was superior to APH in the PR19 models and therefore it preferred to use booster pumping stations. The CMA considered that APH makes sense from an engineering and economic perspective. However, it had concerns regarding the quality of the APH data and its statistical significance. Source: https://assets.publishing.service.gov.uk/media/60702370e90e076f5589bb8f/Final_Report_---_web_version_-_CMA.pdf p.139.

[15] Middle layer Super Output Area (MSOA) estimates are sourced from the Office of National Statistics (ONS) and provide a more granular view of density compared to the Local Authority Distract area data underpinning the PR19 weighted average density measure.

[16] Severn Trent response to Ofwat's consultation on assessing base costs at PR24. Source: https://www.ofwat.gov.uk/wp-content/uploads/2021/12/SVE_response_to_assessing_base_costs_at_PR24.pdf

Cost driver	Explanatory variable	Explanation	Models included
	• *Alternative weighted average treatment complexity variable*	*Water companies report the volume of water treated at treatment works by different complexity levels, ranging from zero to six.* *At PR19, Ofwat used two measures to control for treatment complexity: the percentage of water treated at water treatment works with complexity level 3 or higher, and weighted average complexity.* *Ofwat used the former because it considers that there is a step change in treatment costs between works of complexity level 2 or less and works at higher levels of complexity. Levels 0, 1 and 2 include relatively simple works, such as those treating good quality groundwater sources. Level 3 treatment and upwards involves works utilising multiple treatment stages to treat water from lower quality sources. At PR19, some companies concurred with this view, but some companies suggested level 4 would be a more appropriate cut-off.* *The weighted average complexity driver used by Ofwat at PR19 was calculated based on the proportion of water treated at each complexity level and the numbers one to seven, each corresponding to a treatment complexity level.* *We tested alternative complexity band variables including different cut-offs (bands 4-6, bands 5-6, band 6) and using separate complexity band variables for surface water and groundwater.* *For the weighted average treatment complexity variables, we tested alternative weights proposed by Severn Trent Water which suggested that there is little difference in the cost of treating water between levels 2 and 3, and also between 4 and 5.*[17] *We also tested using squared weights.*	
Alternative scale driver	• *Length of mains*	*At PR19, Ofwat used the number of connected properties as the scale driver in the WW models, consistent with the scale driver used in the WRP models. However, length of mains was used as the scale driver in the TWD model.* *An argument for using length of mains in the WW models is that TWD makes up the largest share of wholesale water base costs.*	• *WW*
Network reinforcement drivers	• *New properties as a percentage of total properties* • *Annual percentage increase in properties* • *Annual population growth*	*Variables such as new properties as % of total properties and the annual % increase in number of properties are potential drivers of network reinforcement costs.* *If network reinforcement costs are included in base costs, scale drivers (e.g., connected properties) in botex models may capture growth costs imperfectly. Companies that experience high growth may be disadvantaged when growth is not explicitly controlled for in the model.* *Our testing was focused on TWD models given that this is where network reinforcement costs are reported.*	• *TWD*

[17] Severn Trent response to Ofwat's consultation on assessing base costs at PR24. Source: https://www.ofwat.gov.uk/wp-content/uploads/2021/12/SVE_response_to_assessing_base_costs_at_PR24.pdf

Cost driver	Explanatory variable	Explanation	Models included
Weather related variables	• *Potential Evapotranspiration (PET)* • *Annual rainfall* • *Peak Distribution Input (DI, 7-day rolling average), and peak DI relative to annual DI*	*Long-term trends in weather related variables, driven by climate change, may result in changes in company costs. For example, depletion of water resources (due to higher temperature and lower rainfall) may affect the quality of raw water and increase treatment requirements.*	• *WRP* • *TWD* • *WW*
Economies of scale in water resources	• *Number of sources / DI per source*	*A higher number of water sources may be associated with higher costs either because the average size of the water source is smaller resulting in diseconomies of scale, or because it may reflect the larger size of a company (i.e., capture scale effects).* *The average size of water sources (DI per number of sources) may be a better measure of economies of scale than number of sources. Obtaining a large proportion of water from a small number of sources may lead to lower unit costs.* *However, a key factor is likely to be the size of treatment works. The cost impact of drawing water from many small sources may be limited if these are all linked to one, or a small number, of larger treatment works.* *There may also be an opposite impact due to the potential link between number of sources and treatment complexity. Smaller water sources tend to be groundwater sources which typically require less treatment (and have lower costs), while larger sources (e.g., rivers) typically require more complex treatment.*	• *WRP*
Economies of scale in water treatment	• *% of DI treated in different treatment size bands*	*Large treatment works are expected to have a lower unit cost of treatment than small treatment works.* *We would expect a positive coefficient for smaller works; the more input treated in smaller works, the higher the cost.*	• *WRP*
Water resource drivers	• *% DI from reservoirs* • *Total number of reservoirs* • *% DI from pumped reservoirs* • *% DI from impounding reservoirs* • *Total number of impounding reservoirs*[18] • *% DI from all pumped sources*	*Different water sources might be expected to have different costs. Pumped water sources are likely to have higher operational costs compared to impounding reservoirs due to the power costs of pumping.* *Reservoirs may also have higher ongoing maintenance costs compared to other sources because of additional inspection and maintenance requirements set out in the Reservoir Act 1975.* *Pumped storage reservoirs are therefore expected to be the most expensive water source to operate and maintain.* *We tested a range of different variables capturing different water sources.*	• *WRP* • *WW*

[18] For total number of impounding reservoirs, we used variable BN4803S from Ofwat's wholesale water dataset.

Cost driver	Explanatory variable	Explanation	Models included
Time trend	• *Time trend*	*A time trend can be used to capture factors such as ongoing efficiency, real price effects, and any other temporal factors affecting the entire sector that are not picked up by the other explanatory variables. As a result, the expected sign of the estimated coefficient on the time trend is ambiguous.* *Some companies argue that a time trend should be included in the wholesale water models to explain the industry wide increase in costs seen in recent years. This may reflect increasing costs incurred to address leakage and water efficiency schemes.* *The use of a time trend is one way of addressing the issues highlighted above. However, it does not address the root cause behind increasing expenditure. Other solutions may include:* *- Controlling for additional cost drivers that explain the increase in expenditure more precisely (e.g., including a leakage variable)* *-Reducing the sample size if there is evidence of a structural break that cannot be explained (recognising this would reduce the sample size and may reduce the precision of model parameter estimate).* *- Calculating a catch-up efficiency adjustment based on more recent data (i.e., Ofwat PR19 approach).*	• *WRP* • *TWD* • *WW*
Wastewater models			
Weather related variables	• *Annual rainfall* • *Urban rainfall* • *Normalised urban rainfall (divided by sewer length)*	*Rainfall affects the volume of inflows into sewerage networks. An increase in the level of peakiness of rainfall strains network and storage assets, requiring larger network and storage assets, and more pumping of sewage to treatment plants.* *The rainfall measures considered act as a proxy for drainage inflows. However, this variable is unable to capture the "peakiness" of drainage flows.* *At PR19, Ofwat tested a variable related to drainage costs in the sewage collection models. This urban runoff variable reflects urbanisation rates, rainfall and soil permeability. While the variable was found to be statistically significant and positive in sewage collection models at PR19, it caused the main scale variable to become insignificant.*[19] *Urban rainfall is derived by multiplying total rainfall and total urban area in each company area (i.e., urban acts as a dummy, multiplying the rainfall driver with the area that is classified as 'urban' by*	• *SWC* • *WWWN+*

[19] Ofwat, PR19 Supplementary Technical Appendix: Econometric approach: https://www.ofwat.gov.uk/wp-content/uploads/2019/02/Supplementary-technical-appendix-Econometric-approach-1.pdf

Cost driver	Explanatory variable	Explanation	Models included
		ONS). It is preferred to the simple annual rainfall measure because it is expected that rainfall has a larger impact on sewage collection costs in dense, urban areas. We tested four urban rainfall measures, which vary by the granularity of areas considered (i.e., using geographical boundaries used in the LAD and MSOA density measures), and by the inclusion or exclusion of soil permeability factors. *The data and statistical results for these four variables are very similar. Therefore, throughout this report, we only consider the urban rainfall variable weighted using LAD, excluding soil permeability. The soil permeability measure has been excluded because of concerns around the availability and transparency of the underlying data.* *The urban rainfall measure is not adjusted for scale (the rainfall measure is multiplied by the size of the urban area). For this reason, we also consider a normalised urban rainfall driver that divides urban rainfall by sewer length, to improve the interpretation of the variable by isolating the rainfall effect from scale effects.*	
Sewage treatment complexity	• *% load with Ammonia consent* • *% load with Biochemical Oxygen Demand (BOD) consent* • *% load with Phosphorus consent* • *% load with UV treatment*	*Higher wastewater treatment complexity can drive higher costs due to increased chemical and energy requirements. More onerous discharge consent limits tend to require more, or larger, treatment process units and are therefore more costly to comply with.* *At PR19, Ofwat used the proportion of load with ammonia consent <=3 mg/l to account for treatment complexity. Alternative drivers exist, such as phosphorus consent, BOD consent, and the usage of UV treatment.* *In DEFRA's May 2022 Consultation on Environmental Targets, it proposed a target to reduce phosphorus levels in wastewater by 2037 against a 2020 baseline. Considering this increased focus on reducing phosphorus, we investigated the relationship between phosphorus treatment and costs.* *We also consider the use of more than one treatment driver, to reflect the possibility that different treatment complexity drivers may not be correlated.*	• *SWT*
Network topography and coastal population	• *% of a company's population in coastal areas* • *Pumping capacity per sewer length*	*Companies that serve coastal areas may incur higher costs due to the need to pump more sewage to and from treatment works. Operating in coastal areas may increase sewage treatment costs due to the need to operate in restricted areas, the requirement for extensive pumping and treatment away from coastal urban areas, more corrosive operating environments, and requirements on the effluent quality discharged offshore to comply with environmental permits.* *This may result in higher pumping costs, although a potential mitigating factor is the fact that coastal areas may be less hilly, which is also a factor affecting pumping costs.* *At PR19, pumping capacity per sewer length was used to capture the impact of energy costs related to pumping capacity for sewage collection, but no equivalent driver was used in the sewage treatment models.*	• *SWC* • *SWT* • *WWWN+*

Cost driver	Explanatory variable	Explanation	Models included
		For the percentage of a company's population in coastal areas, we use information provided by Southern Water based on ONS data. We have not verified or tried to replicate this data.	
Economies of scale in sewage treatment	• *% load treated above different STW size cut-offs* • *Weighted average bands (WAB), using two alternative band systems* • *Weighted average treatment size (WATS)*	*Large treatment works are expected to have a lower unit cost of treatment than small treatment works.* *At PR19, Ofwat used two measures in the sewage treatment and bioresources plus models:* • *the proportion of load treated at small works of in bands 1-3 (≤ 2,000 population), measuring any diseconomies of scale from operating small works; and* • *the proportion of load treated at the largest category of works in band 6 (≥ 25,000 population), to capture economies of scale at large treatment works.* *When we run the PR19 SWT models with the updated PR24 dataset, the band 6 variable is only significant at the 10% level. As noted by the CMA in its PR19 redetermination, size band 6 (≥25,000 population served) includes a large range of treatment works.[20] It is therefore possible that this variable does not capture enough variation between companies.* *We explore a range of alternative economies of scale drivers. The first set is the proportion of load treated in STWs of a range of sizes. We test this for cut-offs ranging from 100,000 to 1,000,000 population, chosen according to the alternative band thresholds described below.* *Additionally, we test two Weighted Average Band (WAB) measures. These weight the band numbers by the load treated in each size band (i.e. 1 x % load treated in Band 1, 2 x % load treated in Band 2, etc.). The WAB is calculated for two alternative band systems that split band 6.* *The first is a band system proposed by CEPA's engineering team:* • *Band 6: 25,000-100,000 population* • *Band 7: 100,000-500,000 population* • *Band 8: 500,000+ population* *The second band system was proposed by Anglian Water:* • *Band 6: 25,000-125,000 population* • *Band 7: 125,000-250,000 population* • *Band 8: 250,000-500,000 population* • *Band 9: 500,000-1,000,000 population* • *Band 10: 1,000,000+ population*	• *SWT* • *WWWN+*

[20] CMA (2021), Anglian Water Services Limited, Bristol Water plc, Northumbrian Water Limited and Yorkshire Water Services Limited price determinations, p. 156, https://assets.publishing.service.gov.uk/media/60702370e90e076f5589bb8f/Final_Report_---_web_version_-_CMA.pdf

Cost driver	Explanatory variable	Explanation	Models included
		Thirdly, we consider a weighted average treatment size (WATS) variable, which uses an Ofwat dataset detailing STWs in the current size band 6. The aim of the WATS variable is to calculate the weighted average size of a sewage treatment works for each company. *For larger STWs in the current size band 6, the WATS variable sums the size of each plant weighted by the percentage of load treated by each plant. For treatment works in size bands 1-5, the calculation takes the average STW size for each band weighted by the percentage of load treated in each of size band 1-5.*[21] *The alternatives considered are all ways of measuring and capturing economies of scale at SWTs and reflect different assumptions regarding the relationship between STW size and unit costs. For example, the STW cut-offs assume that there is a specific threshold where companies achieve economies of scale; this assumption is valid if there is a step-change in unit costs around the assumed cut-off.* *On the other hand, the two WAB and the WATS drivers consider all load and allow costs to decrease with the size of STWs. Compared to the WATS, the WAB variables require additional explicit assumptions to be made regarding the definition and split of bands.* *The true relationship between STW size and unit costs would determine which of these drivers is most applicable. We test a range of alternatives in order to identify a suitable driver, considering technical rationale and statistical performance.*	
Alternative density variables	• *MSOA*[22] • *Local Authority District (LAD) from MSOA*	*Density can have two counteracting effects on costs: on one hand, dense areas may have high costs due to higher labour, property and access costs; on the other hand, dense areas can be served by larger treatment works which incur lower unit costs.* *The alternative density variable (MSOA) uses more granular ONS data than was available at PR19, avoiding issues created by the merging of LADs over the time series covered by the base models. We also tested a density variable that uses MSOA data to create LAD level density (referred to as "LAD from MSOA").*	• *SWC* • *WWWN+*
Network reinforcement drivers	• *New properties as a percentage of total properties*	*As explained for the equivalent water models option, any growth-related costs included in base costs may not be entirely captured by scale drivers.*	• *SWC*

[21] Formulaically, the WATS of company *i* can be shown as: $WATS_i = \Sigma_{j=1}^{5} \left(\frac{load\ treated\ in\ band\ j\ for\ company\ i}{nr\ treatment\ works\ in\ band\ j\ for\ company\ i} \times \frac{load\ treated\ in\ band\ j\ for\ company\ i}{load\ treated\ for\ company\ i}\right) + \Sigma_{k\in j=6}\left(load\ treated\ in\ STW\ k\ for\ company\ i \times \frac{load\ treated\ in\ STW\ k\ for\ company\ i}{load\ treated\ for\ company\ i}\right)$

[22] Middle layer Super Output Area (MSOA) estimates are sourced from the Office of National Statistics (ONS) and provide a more granular view of density compared to the Local Authority District area data underpinning he PR19 weighted average density measure.

Cost driver	Explanatory variable	Explanation	Models included
	• *Annual percentage increase in properties* • *Annual population growth*		
Time trend	• *Time trend*	*As suggested in the description of the time trend for the wholesale water models, a time trend can be used to capture various factors such as ongoing efficiency, real price effects, and any other temporal factors affecting the entire sector that are not picked up by the other explanatory variables and therefore it is difficult to form expectations about the sign of the estimated coefficient on the time trend.* *While the argument for using a time trend to capture increases in costs is less relevant for wastewater, for completeness we also tested a time trend in the wastewater models.*	• *SWC* • *SWT* • *WWWN+*
Bioresources models			
Alternative density variables	• *MSOA* • *Local Authority District (LAD) from MSOA* • *Properties per sewer length* • *Squared density term*	*Density acts as a proxy for two separate cost effects:* • *Economies of scale at sludge treatment level - denser areas may tend to have larger sludge treatment centres that can bring down unit costs.* • *The amount of transportation companies need to undertake in sludge transport and sludge disposal activities.* *PR19 used WAD – LAD measure. We explore alternative density measures, as for the water and wastewater models.* *We also test the inclusion of a squared density relationship. The squared density measure aims to capture increased bioresources costs of extreme density levels (i.e., most dense and sparse areas).*	• *BR1*
Sewage treatment complexity	• *% load with Ammonia consent* • *% load with Phosphorus consent*	*Sewage treatment complexity accounts for the impact of wastewater treatment complexity on the quality of sludge which enters the bioresources value chain.* *We test the impacts of ammonia and phosphorus consents which could make the sludge more costly to treat. At PR19, Ofwat used the proportion of load with treatment work consents with ammonia <=3 mg/l to account for sewage treatment complexity in the BRP models, but treatment complexity drivers were not used in the BR models.* *We also consider the use of more than one treatment driver, to reflect the possibility that different treatment complexity drivers may not be correlated.*	• *BR1* • *BR2*
Economies of scale in sewage treatment works	• *% load treated above different STW size cut-offs* • *Weighted average bands (WAB), using two alternative band systems.*	*The economies of scale drivers tested are specific to sewage treatment plants therefore they may imperfectly capture the impact of economies of scale in bioresources.* *The rationale for testing them in the bioresources models is that the prevalence of small sewage treatment works may imply more sludge has to be transported to central sludge treatment centres.*	• *BR1* • *BR2*

Cost driver	Explanatory variable	Explanation	Models included
	• *Weighted average treatment size (WATS)*	*In contrast, large sewage treatment works may have sludge treatment facilities on-site therefore reducing the costs associated with sludge transport.* *Therefore, these drivers are more likely to capture sludge transport (and sludge disposal costs I rather than sludge treatment costs.* *We test the same variables as for the sewage treatment models.*	
Time trend	• *Time trend*	*For completeness and consistency with our testing for water and wastewater models, we have tested the inclusion of a time trend in the BR models as well.*	• *BR1* • *BR2*

5. PHASE 1 ASSESSMENT

In Phase 1, we carried out a preliminary selection of models based on our analysis of possible alternative variables set out in Table 4.2 above and using the criteria and statistical tests described in Section 2. The purpose of the Phase 1 assessment was to identify promising model specifications that warranted a more extensive evaluation in our Phase 2 assessment.

In Phase 1, we evaluated models against 2 key criteria:

- **Rationale:** We considered the economic and technical rationale of the model, and whether the model was consistent with Ofwat's regulatory policy (e.g., risk of perverse incentives, focus on exogenous cost drivers).
- **Statistical results**: We then considered initial statistical results of the model. This included an evaluation of data availability and quality, as well as basic statistical performance such as predictive power (R-squared) and parameter significance.

We scored each model using a RAG rating:

- ➕ **Progress:** The model demonstrates a clear rationale and strong statistical results. We progressed these models to Phase 2 assessment.
- ➖ **Marginal:** The model demonstrates a clear rationale but provide mixed initial statistical results; or the model displays strong statistical results but has no clear rationale. Models that scored marginally were progressed to Phase 2 assessment, for further analysis.
- ❌ **Do not progress:** The model demonstrates a poor rationale, performs poorly in terms of statistical results or fails to meet regulatory objectives e.g. on drivers being exogenous. We did not progress these models to Phase 2 assessment.

5.1. Assessment of alternative water model variables

We present below a summary of the findings of the assessment in Phase 1 for each of the explanatory variables considered in our analysis for wholesale water models. We present full modelling results in Appendix A.

Table 5.1: Summary of our Phase 1 Assessment for wholesale water models

Model	WRP1	WRP2	TWD	WW1	WW2
Average Pumping Head (APH)					
APH used alone or in place of booster stations	❌	❌	➕	➕	➕
APH used in addition to booster stations			➕	➕	➕
Alternative density drivers					
MSOA (population)	➕	➕	➕	➕	➕
LAD from MSOA (population)	➕	➕	➕	➕	➕
Properties per length of mains	➕	➕	➕	➕	➕
Remove squared density term	❌	❌	❌	❌	❌
Alternative water treatment complexity drivers					

Model	WRP1	WRP2	TWD	WW1	WW2
Alternative treatment complexity bands	−			−	
Alternative weights for weighted average complexity		+			+
Alternative scale drivers in wholesale water models					
Length of mains				+	+
Network reinforcement drivers					
New properties as percentage of total properties			×		
Annual percentage increase in properties			×		
Annual population growth			×		
Weather related drivers					
Annual rainfall	×	×	×	×	×
Potential Evapotranspiration (PET)	×	×	×	×	×
Peak demand	×	×	×	×	×
Economies of scale in water resources					
Number of sources	×	×			
Average DI per source	×	×			
Economies of scale at water treatment works					
Percentage of treatment in band 8	×	×			
Percentage of treatment in base 7-8	×	×			
Percentage of treatment in bands 6-8	×	×			
Percentage of treatment in band 1	×	×			
Percentage in treatment of bands 1-2	×	×			
Percentage of treatment in bands 1-3	×	×			
Water resources drivers					
% DI from reservoirs	−	−		−	−
% DI from pumped reservoirs	−	−		−	−
% DI from impounding reservoirs	×	×		×	×
Total number of pumped reservoirs	×	×		×	×
Total number of impounding reservoirs	×	×		×	×
Total number of reservoirs	×	×		×	×
% DI from all pumped sources	×	×		×	×

Model	WRP1	WRP2	TWD	WW1	WW2
Time trend					
Time trend	✕	✕	–	✕	✕

Key: + *Progress* – *Marginal* ✕ *Do not progress*

Below we discuss in more detail which models were selected for progression to Phase 2, based on our Phase 1 selection criteria.

Average pumping head

APH has a strong engineering rationale as a driver that accounts for pumping requirements, and it produced good statistical results in the wholesale water and treated water distribution models.

Average pumping head (APH) has been used by Ofwat in the past as a variable in its econometric models, however it was not included in the PR19 models partly due to concerns over data quality. Ofwat has taken steps in recent years to improve the data quality for APH, for example by conducting a study into the companies' reporting of APH[23]. This has led to better and more consistent reporting by companies in the last few years. Based on the findings of this study and recent improvements in data reporting, Ofwat considers the APH for treated water distribution to be the largest contributor to wholesale water APH, and is better quality data than other components of APH, such as APH in water treatment. Despite this, some data quality concerns persist. In our initial review of the data, we observed potential anomalies in the APH data in the form of large spikes or drops in the data. Ofwat addressed some of these issues in the updated dataset published in November 2022 however this indicates that some further review of the data may be necessary.

We found that APH (all) in the WW models and APH (treated water distribution) in the TWD model performs well in terms of statistical robustness, both alongside and instead of booster stations per length. The estimated coefficients on the APH variable in these models are statistically significant at the 1% level and the explanatory power of the models is similar to that of the PR19 model specifications. However, we found that coefficients on APH (transport, resources and treatment) are statistically insignificant in the WRP models.

As there are some concerns regarding the quality of APH (WRP) data, and this driver was not statistically significant in the WRP models, in our Phase 2 testing we have also included APH (TWD) instead of APH (all) as a driver in the WW models.

Alternative density drivers

The alternative density drivers were progressed to Phase 2 as there is a good justification and statistical results were not very different relative to the PR19 specifications. The option to remove the squared density term was not pursued, as the rationale for it is not strong, and the statistical results were weak.

The PR19 LAD measure of density requires mapping of company areas to LADs and is sensitive to changes in ONS boundaries. Some of the alternative density measures considered here seek to avoid that issue:

- The MSOA weighted average density (WAD) variable and the MSOA WAD aggregated at the LAD level use more granular data than the LAD measure used at PR19 and do not rely on mapping of companies' area to LADs. However, the LAD from MSOA measure remains sensitive to changes in ONS boundaries.
- The property per length variable is a simpler measure that does not rely on mapping of areas to LAD and is not affected by changes in LAD boundaries. It was used by Ofwat at PR19 in the wastewater models. However, as length of mains is, to some extent, under a company's control, this variable could be regarded as more

[23] Average Pumping Head: data quality improvement report, https://www.ofwat.gov.uk/publication/average-pumping-head-data-quality-improvement-report/

endogenous than the other density measures. As a result of its simplicity, it may also fail to capture differences in density within a company's operating region as well as other measures do.

We find that the MSOA, MSOA-LAD WAD, and properties per length variables are statistically significant and provide a similar level of explanatory power as the PR19 models. We therefore progress all these alternative density variables to Phase 2 as each represents a potentially viable alternative to the PR19 density measure.

The argument for removing the squared density term put forward by Severn Trent was that the original density driver suffered from specification issues.[24]

In particular Severn Trent commented that in TWD models density should only capture higher cost in urban areas, with cost impacts in less dense area better captured by network complexity drivers. Similarly, for WRP models, the company argued that density effects may be better captured through economies of scale drivers. The statistical results of the models excluding the square density term are weak, with a loss in explanatory power in all models, and the density driver becomes statistically insignificant in some models. Similar results are obtained when using alternative MSOA density measures without the squared term.

Alternative water treatment complexity drivers

We decided on balance to progress one of the alternative treatment complexity bands drivers (surface water and groundwater bands 3-6) as it produces statistical results similar to the PR19 drivers. We also progressed the alternative weighted average treatment complexity driver as the coefficient was more statistically significant in the WRP2 model than the PR19 variable.

Based on a review of the data, we found that most companies treat more than 80% of their water in bands 3-6. We tested alternative water treatment complexity band drivers with different band combinations (e.g., water treated in bands 4-6, bands 5-6 or band 6) to potentially capture more variation between companies within higher treatment bands. We found that the coefficients on water treated in bands 4-6 and 5-6 are statistically insignificant in the WRP1 model, while significant in the WW1 model. The coefficient on water treated in band 6 is statistically significant in the WRP1 model but has a counterintuitive negative sign and the variable is not significant in the WW1 model. The R-squared decreases in all models compared to the PR19 model, especially so in the WRP models. As a result, we did not progress any of these options to Phase 2.

As there may be different costs associated with treating surface water sources compared to groundwater sources and almost all surface water is treated in bands 3-6, we also tested separate treatment band variables for surface water and groundwater, as suggested by Severn Trent Water.[25] The models tested included percentage of groundwater treated in bands 3-6 together with percentage of surface water treated in bands 3-6, bands 4-6, bands 5-6 or band 6. We found that the coefficient on groundwater treated in bands 3-6 is statistically significant in all alternative specifications. The coefficient on surface water bands is only statistically significant and has an intuitive sign for bands 3-6. On balance we decided to progress surface water and groundwater bands 3-6 to Phase 2 for further testing although the rationale for splitting treatment complexity into surface water and groundwater would require further justification.

The alternative WAC models based on the complexity levels proposed by Severn Trent Water[26] were statistically significant, and the PR19 WAC variable was not statistically significant in WRP2 at the 10% level (p value = 0.123). Therefore, we progressed these alternative WAC variables (both the linear and squared specifications of the

[24] Severn Trent response to Ofwat's consultation on assessing base costs at PR24. Source: https://www.ofwat.gov.uk/wp-content/uploads/2021/12/SVE_response_to_assessing_base_costs_at_PR24.pdf

[25] Severn Trent response to Ofwat's consultation on assessing base costs at PR24. Source: https://www.ofwat.gov.uk/wp-content/uploads/2021/12/SVE_response_to_assessing_base_costs_at_PR24.pdf

[26] The alternative complexity levels proposed by Severn Trent are 1,2,3,3,4,4,5 reflecting the fact that, in the company's view, there is little difference in the cost of treating water between levels 2 and 3, and also between 4 and 5.

alternative weights) to Phase 2, however we note that the technical rationale for these weights requires further consideration.

Alternative scale driver in wholesale water models

We progressed the alternative scale driver to Phase 2. The rationale for length of mains as a scale driver in WW models is strong given that it is the chosen scale driver for the TWD model, and the statistical results were strong, as explained below.

The coefficients on lengths of mains in the WW models are statistically significant at the 1% level, and of a similar magnitude as “number of properties”. However, the coefficients on number of booster stations per length becomes less significant, at the 10% level, instead of the 1% level in WW1 and 5% level in WW2. The explanatory power of the WW models with length of mains as the scale variable is similar to that of the PR19 model specifications.

Network reinforcement drivers

We did not progress any network reinforcement drivers to Phase 2. While there is an intuitive rationale for these cost drivers, we found that the statistical results were very weak.

Network reinforcement drivers are intended to capture instances where the models may not capture the higher costs incurred by a company that experiences high growth. However, the scope of growth-related costs included in the definition of base costs is reduced compared to PR19 (for example, site-specific growth expenditure is now excluded from base costs). As a result, the network reinforcement costs included in base costs may not be sufficiently material to justify controlling for within the model. We found that the statistical results were very weak, with none of the variables being statistically significant, and the explanatory power of the models being virtually unchanged when including these variables.

Weather related variables

We did not progress any climate change or weather-related drivers to Phase 2. The link between the drivers tested and company costs is unclear especially over a short period of time, and the statistical results are weak for most models.

Weather related factors may affect water resources or the demand for water. It is unclear however how strong the link is between weather-related variables and base costs. These drivers may affect companies gradually over the long term, but may not explain variation in costs between companies given that all companies are expected to be affected by the same trends.

The annual rainfall variable is statistically insignificant in all models. Although the PET variable is statistically significant in the TWD model, we considered that the overall justification for this cost driver is weak as it does not perform well in any of the other models. The PET variable capture temperature and other environmental factors that may result in higher water evaporation. While we may expect this to affect water resources and water treatment, it is less clear how this would impact TWD base costs.

While peak DI is statistically significant in the TWD model, we think this is likely to capture scale effects as peak DI and length of mains are highly correlated. This is also indicated by the fact that the magnitude of the coefficient on the scale driver in the TWD model (length of mains) reduces by an amount similar to the coefficient on peak DI. When normalising peak DI by annual DI, we find that the performance of the model (only marginally significant at the 10% level in the WW1 model) is not strong enough to justify further consideration.

Economies of scale in water resources

We did not progress the economies of scale in water resources drivers. The link between the proposed economies of scale drivers and base costs is unclear, as explained below, and the statistical results were weak.

Obtaining a large proportion of water from a small number of sources may lead to lower unit costs. However, a key factor is likely to be the size of treatment works. The cost impact of drawing water from a large number of small sources may be limited if these are all linked to one, or a small number, of larger treatment works. There may also be an opposite impact due to the potential link between number of sources and treatment complexity. Smaller water sources tend to be groundwater sources which typically require less treatment (and thus less costs), while larger sources (e.g., rivers) typically require more complex treatment.

We found that both number of sources and DI per source are statistically insignificant in the WRP models.

Economies of scale in water treatment

We did not progress any economies of scale at water treatment works variables to Phase 2. The materiality of economies of scale is less clear than in WWW models, and the statistical results were counterintuitive.

Given that large treatment works are expected to have a lower unit cost of treatment than small treatment works (and vice versa), we expected a positive coefficient for smaller works, and a negative coefficient for larger works.

The statistical results were either insignificant, or with a coefficient with a counter-intuitive sign, suggesting dis-economies of scale. Ofwat also informed us of potential data issues due to changes in how the size of a water treatment works is defined in Regulatory Reporting Guidelines, which is likely to affect comparability over time. Therefore, we did not take these models further to Phase 2

Water resources drivers

On balance we decided to progress the % of DI from (pumped) reservoirs to Phase 2 for further testing as the statistical results were reasonably strong.

The technical rationale for cost drivers reflecting the water input sources suggested that pumped water sources may be more expensive given additional pumping costs while reservoirs might also be expected to be more expensive given additional inspection and maintenance costs.

We tested a range of different variables to test these impacts. We found that % of DI from reservoirs or pumped reservoirs is statistically significant in the WRP and WW models, even though the R-squared decrease slightly in the WRP models. However, in the case of % of DI from pumped reservoirs, the density variables in both WRP models lose significance and/or the magnitude of the estimated coefficients reduces in size.

Other alternative specifications tested included:

- % DI from all pumped sources
- % DI from impounding reservoirs
- Total number of pumped, impounding or total reservoirs.

We considered that none of these alternative specifications performed sufficiently well to progress to Phase 2. In some cases, the coefficients were either statistically insignificant (e.g., number of pumped storage reservoirs) or had a counterintuitive sign (e.g., % DI from pumped storage), and the explanatory power of the models was virtually unchanged or decreased slightly. The total number of reservoirs was statistically significant at the 10% level but also resulted in the economies of scale driver becoming insignificant and the R squared reducing slightly. In addition, we understand that there are potential data quality concerns with this variable due to changes in how the number of reservoirs has been reported over time.

Time trend

On balance we progressed the linear time trend variable in TWD model to Phase 2 assessment as the time trend is statistically significant in this model.

As discussed in section 4.2, a linear time trend may capture a range of factors that affect costs over time, such as technological change and real price effects. It may also pick up other factors that have potentially led to greater costs over time, such as the costs associated with leakage reduction or climate change. However, such changes over time might also be captured by other variables. Overall, there is no clear expected sign on the time trend variable as the time trend may pick up different factors that have opposite impacts on costs.

We found that the linear time trend is only statistically significant in the TWD model. The explanatory power increased marginally in the TWD model, but decreased in all other models. As the time trend coefficient is positive, including a time trend in the regression would allow costs to gradually increase over time. The rationale for allowing costs to gradually increase over time requires further consideration. Ofwat could consider other solutions for controlling trends in expenditure over time such as including additional cost drivers that capture the root cause of the cost increase.

5.2. Assessment of alternative wastewater model variables

We present a summary of the findings of the assessment in Phase 1 for each of the cost drivers considered in our analysis, in Table 5.2 below. We present full modelling results in Appendix A.2.

Table 5.2: Summary of our Phase 1 Assessment for wholesale wastewater models

Model	SWC1	SWC2	SWT1	SWT2	WWWN+
Baseline WWWN+ model					
WWWN+ model					+
Rainfall					
Annual rainfall	×	×			×
Urban rainfall	−	−			−
Urban rainfall per sewer length (log)	−	−			−
Sewage treatment complexity					
Ammonia consents ≤ 1mg/l (%)			×	×	
Phosphorus consents ≤ 0.5mg/l (%)			×	×	
Phosphorus consents ≤ 1mg/l (%)			×	×	
BOD consents ≤ 7mg/l (%)			−	−	
BOD consents ≤ 10mg/l (%)			×	×	
Load treated with UV ≥ 30mW/s/cm^2 (%)			×	×	
Composite variables			×	×	
Network topography and coastal population					
'Coastal' population (%)	×	×	−	−	×
Pumping capacity per sewer length			−	−	
Sewage treatment economies of scale					
Anglian Water size bands				+	+
CEPA size bands				+	+
Weighted average bands				+	+
Weighted average treatment size				+	+
Density					
MSOA density		+			×
MSOA density, aggregated at LAD level		+			×
Network reinforcement					
New properties as percentage of total properties	×	×			
Annual percentage increase in properties	×	×			
Annual population growth	×	×			
Time trend					

Model	SWC1	SWC2	SWT1	SWT2	WWWN+
Time trend	✕	✕	✕	✕	✕

Key: *+ Progress* *– Marginal* *✕ Do not progress*

Below we discuss in more detail the models' performance based on our Phase 1 criteria.

Wastewater network plus models

Overall, the WWWN+ model with load (as scale driver), pumping capacity per sewer length (as network complexity driver) and ammonia consent < 3mg/l (as treatment complexity driver) provides the most robust results. Full model results can be found in Appendix A.2.1.

For the development of wastewater network plus (WWWN+) models, we considered the economic and engineering logic for selecting cost drivers and assessed the models based on our selection criteria. We aimed to combine a set of drivers that control for the principal cost drivers (scale, treatment complexity, density and economies of scale).

We present the results of 16 explanatory WWWN+ models, and 5 shortlisted models in Appendix A.2.1. We use these models to assess the relative strengths of potential drivers for the baseline WWWN+ model. Models 1-8 test a combination of the PR19 SWC and SWT models, and Models 9-16 test one or two drivers. Models 17-21 are shortlisted models based on our 16 exploratory models.

We summarise our results by the type of cost driver below. We reference models using the numbering in Appendix A.2.1.

Scale

Load is the most suitable scale driver for WWWN+ models. We find that scale should be a main driver of the WWWN+ models.

The PR19 wastewater models had different scale drivers for SWC (sewer length) and SWT (load). We tested the WWWN+ model with these drivers in addition to properties, to test which driver explains most variation in WWWN+ costs.

We found that all three scale drivers are significant predictors of WWWN+ costs, across all exploratory models. This reflects the importance of the scale driver in costs, in line with expectations. For this reason, we have chosen scale as a main driver in the WWWN+ models.

We conducted univariate regressions of WWWN+ costs on the three scale drivers (WWWN+ models 1 to 3). We found that load provides the greatest explanatory power; this aligns with the observation that 55% of WWWN+ costs can be attributed to Sewage Treatment. The finding of a larger R-squared result for models using load, compared to models using the other two scale drivers, is consistent across all of our exploratory modelling.

Economies of scale

The two STW economies of scale drivers used in PR19 do not perform well in our exploratory WWWN+ models. We further test these and alternative economies of scale drivers as additions to the selected baseline model in Section 6.2.

At PR19, Ofwat used load treated in bands 1-3 and load treated in band 6 as its two drivers for economies of scale in its SWT modelling.

We tested the economies of scale variables in combination with load as a main cost driver (WWWN+ models 4-5). We find that both economies of scale drivers are not statistically significant, though have the predicted sign. In WWWN+ model 4, load treated in size bands 1 to 3 has a p-value of 0.152, and in WWWN+ model 5 load treated in size band 6 has a p-value of 0.549. The R-squared is comparable to the univariate regression model with load, suggesting that the use of an economies of scale driver does not improve the explanatory power of the model.

Across the other exploratory models, the economies of scale drivers are overall statistically insignificant, and in some instances are of the incorrect sign. For example, in WWWN+ model 16, load treated in band 6 is significant at the 1% level (p=0.001) with a positive point estimate of 0.01; this implies that companies with a greater proportion of its load in large STWs face higher costs, contrary to expectations *a priori*.

We note that the STW economies of scale drivers in Models 19 and 20 demonstrate results in line with expectations and provide a strong explanatory power (R^2= 0.952 and 0.949 for Models 19 and 20 respectively). This suggests that the drivers may be useful in the WWNW+ base cost modelling suite. We further test these and alternative economies of scale drivers in Section 6.2 below.

Density

Properties per sewer length appears to be the most suitable density driver for WWWN+ models. However even this density driver is not statistically significant across most model specifications.

At PR19, Ofwat used properties per sewer length and weighted average density in its SWC models. In its PR19 redetermination, the CMA included a squared term for weighted average density in SWC2.[27]

We tested the different density variables in combination with load as a main cost driver (WWWN+ models 6 to 8). We found that properties per sewer length was the only significant density driver, with an expected positive relationship with costs. However, the R-squared value of all three models is slightly lower than the univariate regression model with just load.

Across the other exploratory models, properties per sewer length is consistently of the predicted sign, but is only significant in Models 6 and 13.

It is worth noting that the additional costs of operating a complex network in urban areas may be partially offset by the economies of scale of operating large STWs. Indeed, in WWWN+ Models 13-16 where we control for sewer length and STW economies of scale, the density drivers are of the expected sign and have low p-values.

Network complexity

Pumping capacity per sewer length is a suitable driver for network complexity costs in the WWWN+ models.

At PR19, Ofwat used pumping capacity per sewer length as a driver of network complexity in its SWC models. We tested the single network complexity driver across a range of exploratory models (WWWN+ models 9 to 16). We found that the network complexity driver was statistically significant and of the predicted sign in all eight models.

Treatment complexity

Ammonia consents ≤ 3mg/l is a suitable driver for treatment complexity costs in the WWWN+ models.

At PR19, Ofwat used ammonia consents ≤ 3mg/l as a driver of treatment complexity in its SWT models. We tested the treatment complexity driver across a range of exploratory models (WWWN+ models 9 to 16). We found that it was statistically significant and of the predicted sign in all eight models; the point estimate remains robust across most model specifications.

Shortlisted WWWN+ models

In Appendix A.2.1, we present a shortlist of 5 WWWN+ models (labelled Models 17-21), based on the results of the exploratory models above. These models use load, pumping capacity and ammonia consents as their foundation, and the best performing drivers from density and economies of scale.

Overall, WWWN+ model 21 which uses load, pumping capacity and ammonia consents as the cost drivers provides the most robust results. The 3 cost drivers are all significant at the 1% level, and the point estimates of

---

[27] CMA (2021), Anglian Water Services Limited, Bristol Water plc, Northumbrian Water Limited and Yorkshire Water Services Limited price determinations

each are consistent across model specifications. Furthermore, WWWN+ model 21 demonstrates strong explanatory power (R^2 = 0.943).

We considered the WWWN+ model 21 as the baseline WWWN+ model for Phase 1, where we test the addition of alternative explanatory variables.

Rainfall

We progressed the normalised urban rainfall driver to Phase 2 assessment for the WWWN+ and SWC models. The rationale underpinning the relationship between rainfall and costs is strong, and the normalised urban rainfall driver demonstrates strong statistical results, both in terms of statistical significance of the estimated coefficients and improvements in the explanatory power of the models.

Higher rainfall can affect sewage collection costs, especially in dense, urban areas where large amounts of impermeable surfaces increase the speed of flows of rainfall into sewers. Furthermore, companies in areas of high rainfall may need to spend more to reduce the risks of sewer flooding, which is included in modelled base costs.[28]

There is a concern regarding the relevance of the rainfall driver; the main driver of the effects of rainfall on costs as described above is likely to be rainfall intensity (e.g., mm/hr) rather than annualised rainfall measures. The impacts of rainfall on sewage collection costs are also likely to be very localised and therefore would require very granular data.

Despite concerns with regard to the relevance of the rainfall drivers available, the range of rainfall drivers (annual rainfall, urban rainfall and normalised urban rainfall) demonstrate strong statistical significance as drivers of SWC and WWWN+ costs (statistically significant at or marginally above the 1% level), and improve the explanatory power of the SWC and WWWN+ models, relative to the PR19 specification.

We decided to progress only the normalised urban rainfall driver, as this has the strongest technical rationale. The un-normalised and normalised rainfall drivers perform similarly well based on an econometric assessment of model results. However, the un-normalised urban rainfall driver is affected by scale. This can be seen by the point estimates of the scale driver in the models ran; for example, the point estimate of the sewer length driver falls to 0.725 in SWC Model 1 with the un-normalised rainfall driver, compared with 0.842 in SWC Model 2 where the normalised driver is used. The latter is more comparable to the baseline point estimate of 0.804. By removing the scale effect, the normalised urban rainfall driver also helps with the interpretation of the coefficient. For these reasons, we progress the normalised urban rainfall driver over the un-normalised driver.

Treatment complexity

We progressed the BOD driver to Phase 2, for the SWT models, as it demonstrates a marginal technical rationale and strong statistical results.

There is a strong technical rationale for treatment complexity drivers in general, as higher wastewater treatment complexity can drive higher costs due to increased resource requirements. In particular, DEFRA have highlighted the importance of reducing phosphorus levels in wastewater in its May 2022 Consultation on Environmental Targets. Therefore, we acknowledge that phosphorus removal may become an increasingly important driver of sewage treatment costs. However, there is a lack of relevant historical data on phosphorus consents given that, for most companies, the percentage of load covered by tighter phosphorus limits has only started increasing in recent years. Modelling results also show that phosphorus is not a statistically significant driver of sewage treatment costs. This is likely due to this lack of historical data but may also potentially be due to be interactions between treatment complexity drivers (including phosphorus) and other drivers in the sewage treatment models, such as economies of scale.

[28] Arup recommend testing a driver for drainage of surface water, for which rainfall acts as a proxy. See Arup (2022), Assessment of growth-related costs at PR24.

Of the alternative treatment complexity drivers considered, load treated with BOD ≤ 7mg/l (%) performs the strongest statistically; the variable is statistically significant at the 1% level in both SWT models. However, we have two concerns with this driver:

- The materiality of the impact of BOD on sewage treatment costs is unclear.
- The BOD consents are concentrated within TMS, SVH, NWT and ANH, which account for 84.5% of all load treated with BOD ≤ 7mg/l. It is possible therefore that this driver is picking up company effects that are not related to BOD treatment costs.

Overall, we progress this driver to Phase 2 for further sensitivity testing.

Coastal population

We progressed the coastal population driver to Phase 2 assessment for the SWT models only. The driver demonstrates a marginal technical rationale and the dataset underpinning the variable has not been verified by Ofwat, but it demonstrates strong statistical performance in the sewage treatment model. The pumping capacity per network length may capture the same variation as the coastal driver, but is verified by Ofwat and was included in Ofwat's PR19 sewage collection models. So, we also progress this variable to Phase 2.

Operation in coastal areas is considered to increase sewage collection costs and treatment due to a greater need to pump sewage to and from treatment works. However, the materiality of these impacts is uncertain. Furthermore, the dataset for coastal populations was provided by Southern Water, and has not yet been verified by Ofwat.

We did not progress the coastal variable to Phase 2 for the sewage collection or wholesale wastewater network plus models, as the driver did not improve the performance of the sewage collection models relative to the PR19 specifications, and the coastal variable was not statistically significant. This may be because the coastal effect is somewhat explained by the pumping capacity variable.

However, for the sewage treatment models, the coastal driver substantially improves the models' explanatory power, and the variable was statistically significant in both sewage treatment models (p=0.025 for SWT1 Model 11, and p=0.003 in SWT2 Model 11). Therefore, on balance we think that the coastal variable requires further assessment in Phase 2.

An alternative to the coastal driver is the use of pumping capacity per sewer length. The coastal variable is intended to capture, among other things, effects including greater pumping requirements in coastal areas, which can be measured using the pumping capacity variable. In the dataset used, the correlation coefficient between pumping capacity and the coastal driver is 0.94. Furthermore, the variable was used in the PR19 sewage collection models, and is being used in the baseline WWWN+ model that we developed. However, one issue that Ofwat would need to explore further is whether this correlation between the coastal driver and pumping capacity per sewer length may be, to some extent, spurious (e.g., related to the fact that SRN has extensive coastal populations and a large pumping capacity). This is because we understand that the definition of pumping capacity used by Ofwat excludes pumping to sea outfalls, which could mean that pumping requirements in coastal areas may not be fully captured by the pumping capacity driver used here.[29]

When included in the sewage treatment models, the variable has a low significance (p=0.077 in SWT1 Model 12, and p=0.113 in SWT2 Model 12), but has a positive point estimate as expected. The variable improves model R^2 relative to the baseline, similarly to the coastal driver. Overall, despite its marginal statistical performance, we progress this variable to Phase 2 assessment, because of its potential to capture the impacts of operating in coastal areas on costs and the improved data availability and transparency, relative to the coastal population driver.

[29] RAG 4.11 – Guideline for the table definitions in the annual performance report Draft for consultation, Line 7C.3.

Sewage treatment economies of scale

We progressed the economies of scale drivers to Phase 2 assessment for the WWWN+ and SWT models. The alternative economies of scale drivers based on STW size cut-offs, Weighted Average Bands (WABs) using alternative size bands and the Weighted Average Treatment Size (WATS) driver demonstrate strong technical rationale and statistical results.

The use of economies of scale drivers in general is justified on economic grounds, and the alternative bands improve on the existing band system by disaggregating the current band 6 which contains significant variation in STW size, as noted by the CMA in its PR19 redetermination.[30] Evidence presented by Anglian Water demonstrates the materiality of the effect of economies of scale on unit cost.[31]

The economies of scale drivers we considered were statistically significant and improved the explanatory power of the sewage treatment model, compared to the PR19 baseline. This includes the different STW size cut-offs, the two WAB variables based on the two alternative size bands proposed, and the WATS variable.

Alternative density driver

We progressed the two alternative density drivers to Phase 2 assessment for the SWC models. The drivers demonstrate strong technical rationale, and moderately strong statistical results.

Density was used as a driver of sewage collection costs in PR19, and so its technical rationale is already established. The MSOA and the LAD from MSOA variables rely on more granular data than the LAD driver and avoid the need to rely on mapping of companies' area to LADs.

We considered the effects of replacing the LAD variable with either the MSOA or LAD from MSOA variables, in the SWC model, and adding either alternative of the density variable to the WWWN+ model. SWC2 Model 6 introduces the linear and quadratic LAD from MSOA density driver to the baseline PR19 SWC2 model, replacing the LAD density drivers. The variable performs similarly to the LAD drivers, with similar point estimates, statistical significance and model explanatory power. SWC2 Model 8 replaces the linear and quadratic LAD drivers with MSOA drivers. The MSOA drivers are significant and the models' explanatory power is similar to the baseline, but, differences in the underlying data means that the point estimate of the MSOA driver differs substantially from the PR19 LAD measure.[32]

However, in the WWWN+ model, the linear and quadratic density drivers were statistically insignificant when added to the baseline model. WWWN+ Model 35 includes the LAD density drivers to the model, with p-values of 0.967 and 0.811 for the linear and quadratic LAD variables, respectively. WWWN+ Models 37 and 39 include the linear and quadratic variables for LAD from MSOA and MSOA respectively, finding similar results. We therefore do not progress these models to Phase 2 assessment.

In the three models where the linear term only is included (WWWN+ models 34, 36 and 38), the density drivers are statistically significant (p=0.000 in all 3 models). However, the sign on the density drivers is negative implying costs decrease in more dense areas. This would suggest that the effect of economies of scale (particularly at sewage treatments works) outweighs the impact of more complex operation in dense, urban areas. If this is the case, the economies of scale impacts may be better captured by the inclusion of an economies of scale driver. We also consider that the technical rationale would support the inclusion of a quadratic term in the WWWN+ models due to

---

[30] CMA (2021), Anglian Water Services Limited, Bristol Water plc, Northumbrian Water Limited and Yorkshire Water Services Limited price determinations, Paragraph 4.155

[31] Ofwat Growth Cost Assessment Working Group, November 2021 meeting. The meeting slides can be accessed at: https://www.ofwat.gov.uk/wp-content/uploads/2021/12/CAC-CAWG_slides-11Nov21.pdf

[32] The point estimates of the linear and quadratic LAD drivers in the PR19 SWC2 model are -2.480 and 0.181 respectively; in comparison, the point estimates of the MSOA drivers in SWC2 Model 8 are -5.051 and 0.336, respectively..

the opposing effects of economies of scale at STWs and higher operational costs in urban areas. We therefore do not progress the linear density models to Phase 2.

Network reinforcement drivers

We did not progress the three network reinforcement drivers to Phase 2 assessment. While there is a technical rationale for a positive relationship between network reinforcement and costs, network reinforcement costs included in base costs may not be sufficiently material to control for this within the model. The statistical results of the growth variables are poor, with a significant but counterintuitive sign.

The three network reinforcement drivers have a clear technical rationale; as explained in Section 4.2, scale drivers may capture network reinforcement costs imperfectly. However, it is unclear whether network reinforcement costs are sufficiently material, especially as growth related costs for sewage treatment works are not included in Ofwat's base cost definition.

In the SWC models, the addition of the network reinforcement drivers does not improve the explanatory power of either the SWC models, as measured by the R-squared. Moreover, the coefficients on the three drivers are all statistically significant at the 10% level and negative, suggesting that companies facing higher property or housing growth faced lower costs, contrary to predictions *a priori*. For example, in SWC1 Model 4, the coefficient on New properties as a % of total properties is -0.087 (p=0.082), suggesting that companies with a higher property growth rates face lower costs, all else being equal.

Time trend

We did not progress the linear time trend variable to Phase 2 assessment. The point estimate of the time trend in the models estimated is positive, and the statistical performance of these models was not strong enough to justify further assessment.

We ran all the baseline models with a linear time trend included. Of the five baseline models, the time trend was significant at the 10% level in the two SWT models. In all models, the estimated coefficient of the time trend was zero or positive. The inclusion of a time trend did not substantially improve the explanatory power of the models when compared to the PR19 model specification and the models also failed the RESET test when a linear time trend was included.

CEPA economics matters

5.3. Assessment of alternative bioresources model variables

We present a summary of the findings of the assessment in Phase 1 for each of the cost drivers considered in our analysis, in Table 5.3 below. We present full modelling results in Appendix 0.

Table 5.3: Summary of our Phase 1 assessment for bioresources models

Model	BR1	BR2
Alternative density drivers		
MSOA (population)	– (Marginal)	
LAD from MSOA (population)	– (Marginal)	
Properties per length of mains	– (Marginal)	
Squared density term	✕ (Do not progress)	
Economies of scale in sewage treatment		
Anglian Water's suggested size bands	✕ (Do not progress)	✕ (Do not progress)
CEPA alternative size bands	– (Marginal)	– (Marginal)
Weighted average bands (WAB)	+ (Progress)	+ (Progress)
Weighted average treatment size (WATS)	✕ (Do not progress)	✕ (Do not progress)
Treatment complexity		
Ammonia consents ≤ 1mg/l (%)	✕ (Do not progress)	✕ (Do not progress)
Ammonia consents ≤ 3mg/l (%)	✕ (Do not progress)	✕ (Do not progress)
Phosphorus consents ≤ 0.5mg/l (%)	✕ (Do not progress)	✕ (Do not progress)
Phosphorus consents ≤ 1mg/l (%)	✕ (Do not progress)	✕ (Do not progress)
Composite variables	✕ (Do not progress)	✕ (Do not progress)
Time trend		
Time trend	✕ (Do not progress)	✕ (Do not progress)

Key: + *Progress* – *Marginal* ✕ *Do not progress*

Below we discuss in more detail the models' performance based on our Phase 1 criteria.

Alternative density drivers

While the statistical results are not always strong, the technical rationale for including density is. We therefore progress most of the alternative density variables to Phase 2 assessment.

Density acts as a proxy for two separate costs:

- Economies of scale at sludge treatment level - denser areas may tend to have larger sludge treatment centres that can bring down unit costs.
- The amount of transportation required to undertake in sludge transport and sludge disposal activities.

At PR19, Ofwat used the WAD – LAD measure. We find that the PR19 density driver and all alternative density drivers are statistically insignificant in the BR1 models. We considered that this may be due to a correlation with the economies of scale driver (i.e., density acts as a proxy for economies of scale at sludge treatment level and is therefore highly correlated).

We tested including density with only a scale driver, as we observed that density and EoS drivers are partly capturing the same impact. However, we found that the density variables are not statistically significant even when used in isolation with the scale driver. Including a density variable increases the R-squared only slightly compared to using only the scale variable.

However, on balance, we decided to progress all these alternative density variables to Phase 2 given the strong engineering rationale for including density in the bioresources models and because the alternative density variables represent potentially viable alternatives to the PR19 density measure.

We also tested the inclusion of a squared density relationship. The squared density measure aims to capture increased bioresources costs arising out of extreme density levels (i.e., most dense and sparse areas). However, the coefficients were all statistically insignificant, and the explanatory power of the models did not improve compared to the PR19 model specification.

Economies of scale in sewage treatment

In addition to the PR19 economies of scale drivers, the EoS drivers that we progress to Phase 2 are: % load treated in SWTs serving more than 250,000 population equivalent and SWT WABs variables.

The economies of scale drivers tested are specific to sewage treatment plants. These drivers act as a proxy for economies of scale in sludge treatment given that large STWs are more likely to have co-located sludge treatment facilities and the presence of large STWs may be correlated with the presence of large biotreatment assets. However, these drivers also serve to capture sludge transport and sludge disposal costs given that the prevalence of small sewage treatment works may imply more sludge has to be transported to central sludge treatment centres. In contrast, large sewage treatment works may have sludge treatment facilities on-site therefore reducing the costs associated with sludge transport. In this sense, economies of scale drivers at STWs can be seen to capture similar impacts to the density drivers used in the bioresources models.

The PR19 BR1 model driver (load treated in STWs bands 1-3) generally performs well in most models and has a positive sign as expected, suggesting that the presence of small STWs increases bioresources costs. The PR19 economies of scale driver in BR2 (sewage treatment works per property) is not statistically significant (p value = 0.174). We tested alternative STW economies of scale drivers, particularly focusing on the impact of larger STWs and we found that there is significant variation in terms of the statistical performance of these drivers.

Most alternative economies of scale drivers are statistically insignificant, including the coefficient on the WATS. The exceptions are:

- >250k is statistically significant on a 10% basis;
- the two WABs variables are statistically significant at a 10% level; and
- the % load treated in STWs larger than 1m PE is statistically significant but has a positive sign, which is counterintuitive.

The R-squared decreases in all models using the alternative EoS drivers when compared to the PR19 specification.

As a sensitivity, we also tested a combination of two economies of scale drivers in the BR1 model to capture EoS in higher bands, and diseconomies of scale at lower bands. We found that load treated in STWs bands 1-3 performs well, but the drivers for EoS in higher bands are generally statistically insignificant.

Treatment complexity

We did not progress any treatment complexity variables to Phase 2 assessment. STW treatment complexity may not be a good proxy for the complexity of sludge treatment processes, and the statistical performance of the models does not justify further assessment.

Treatment complexity drivers account for the impact of wastewater treatment complexity on the quality of sludge which enters the bioresources value chain. However, there are likely to be other factors that may have a bigger influence on sludge treatment processes and costs.

Furthermore, we found that most of the drivers tested were statistically insignificant and had a counterintuitive sign. The variables that were statistically significant were:

- % load with ammonia consent below 3mg/l; and
- a composite variable including % load with ammonia consent below 3mg/l and % load with phosphorus consent below 0.5mg/l.

However, for these models, the treatment complexity driver had a negative sign whereas our a priori expectation was that tighter ammonia and phosphorus consents would have a positive impact on bioresources costs.

Time trend

We did not progress the linear time trend variable to Phase 2 assessment. Bioresources costs are not increasing over time, and the statistical performance of these models does not justify further assessment.

For consistency with the water and wastewater models tested, we considered the inclusion of a time trend in the bioresources models as well. We find that the coefficient on time trend is statistically insignificant in all models. Furthermore, the inclusion of a time trend did not substantially improve the explanatory power of the models when compared to the PR19 model specification.

6. PHASE 2 ASSESSMENT

In Phase 2, we further tested the robustness of the model specifications that passed our initial Phase 1 testing, using a range of statistical tests and sensitivity analysis, to identify options that could improve the performance of Ofwat's base cost models relative to the PR19 cost modelling suite.

When assessing the performance of models shortlisted for Phase 2, we considered the following factors (set out in more detail in Table 2.1 and Table 2.2):

- The **explanatory power (R^2)** and **parameter significance (t-test)**.
- Other statistical tests such as **RESET**, **VIF**, **normality** and **pooling** tests.
- **Sensitivity analysis** by excluding years or companies from the sample.

In addition, we also report the **range and stability of efficiency scores** relative to the PR19 model specifications. As we indicate earlier, large changes in a company's efficiency score relative to the PR19 model specifications are not an indication that an alternative model performs better or worse than the PR19 specifications. Rather, it provides an indication of the impacts of selecting alternative models. If the overall range of efficiency scores is larger than the comparative PR19 model specification, however, this could indicate the presence of issues in the underlying model.

In line with our overall approach of building on the PR19 wholesale base cost models, we compare the performance of alternative models to the PR19 model specifications and, given that the PR19 models generally perform well, apply a high hurdle for recommending alternative options as potential additions or replacements to the PR19 model suite. For example, in order to consider recommending a change to Ofwat, we would generally expect a new model to:

- pass the high and medium importance statistical tests set out in Table 2.1;
- prove robust under the range of sensitivity tests applied; and
- meet the wider selection criteria established in Figure 2.1 above.

6.1. WATER MODELS

Table 6.1 summarises the results of our assessment of the variables progressed to Phase 2, based on our findings in Phase 1 as shown in Table 5.1.

Table 6.1: Variables considered in our Phase 2 assessment for the water models

Model	WRP1	WRP2	TWD	WW1	WW2
Average Pumping Head (APH)[33]					
APH (TWD) used alone or in place of booster stations			+	+	+
APH (TWD) used in addition to booster stations			+	+	+
APH (all) used alone or in place of booster stations				–	–
APH (all) used in addition to booster stations				–	–
Alternative density drivers					

[33] We have used APH (treated water distribution) for the TWD and models, and APH (all, i.e., resources, transport, treatment and treated water distribution) for the WW models.

Model	WRP1	WRP2	TWD	WW1	WW2
MSOA	+	+	+	+	+
LAD from MSOA	+	+	+	+	+
Properties per length of mains	+	+	+	+	+
Alternative water treatment complexity drivers					
Alternative treatment complexity driver (SW and GW 3-6 bands	×			×	
Alternative weights for weighted average complexity		–			–
Alternative scale drivers in wholesale water models					
Length of mains				+	+
Water resources drivers					
% DI from reservoirs	×	×		×	×
% DI from pumped reservoirs	×	×		×	×
Time trend					
Time trend			–		

Key: + *Consider* – *Needs further consideration* × *Do not consider*

6.1.1. Water resources plus models

We present the full statistical results of our Phase 2 assessment for the water resources plus models in Table 6.2 and Table 6.3.

Table 6.2: Phase 2 assessment results for the WRP1 models

Driver	WRP1 Model 1	WRP1 Model 3	WRP1 Model 4	WRP1 Model 5	WRP1 Model 24	WRP1 Model 28	WRP1 Model 27
	PR19	MSOA instead of LAD density	LAD from MSOA density	Prop. per length instead of LAD density	Separate SW / GW	% DI from pumped reservoirs	% DI from reservoirs
Properties (log)	1.074***	1.054***	1.077***	1.028***	1.097***	1.076***	1.037***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.006***	0.004***	0.005***	0.005***		0.005***	0.004***
	{0.000}	{0.009}	{0.002}	{0.001}		{0.005}	{0.004}
Density variable (log)[34]	-1.614***	-4.986**	-1.545***	-7.815**	-1.601***	-0.925*	-1.259***
	{0.000}	{0.017}	{0.007}	{0.019}	{0.001}	{0.073}	{0.003}
Squared density variable (log)	0.101***	0.303**	0.097***	0.858**	0.100***	0.05	0.077***
	{0.000}	{0.017}	{0.008}	{0.028}	{0.003}	{0.160}	{0.005}
Surface water treated in bands 3-6 (%)					0.006***		
					{0.003}		
Ground water treated in bands 3-6 (%)					0.003**		
					{0.025}		
DI from pumped reservoirs (%)						0.006***	
						{0.006}	
DI from reservoirs (%)							0.005***
							{0.001}
Constant	-5.093***	9.416	-5.335***	6.988	-5.744***	-7.415***	-5.936***
	{0.000}	{0.226}	{0.000}	{0.309}	{0.000}	{0.000}	{0.000}

[34] The models use the PR19 LAD density variable unless otherwise specified in the column header.

Driver	WRP1 Model 1	WRP1 Model 3	WRP1 Model 4	WRP1 Model 5	WRP1 Model 24	WRP1 Model 28	WRP1 Model 27
R-squared	0.917	0.901	0.909	0.91	0.905	0.906	0.909
RESET test	0.439	0.765	0.436	0.324	0.55	0.733	0.519
VIF (max)*	1.174	1.269	1.206	1.112	1.217	1.185	1.229
Pooling	0.995	1.000	0.999	0.983	0.999	0.983	0.982
Normality	0.128	0.417	0.522	0.143	0.285	0.189	0.096
Heteroskedasticity	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Minimum efficiency score	0.531	0.493	0.527	0.507	0.529	0.501	0.487
Maximum efficiency score	2.022	1.997	2.016	1.975	1.957	2.211	2.346
Sensitivity of estimated coefficients to removal of most and least efficient company	Green	Green	Green	Green	Green	Green	Green
Sensitivity of estimated coefficients to removal of first and last year of sample	Green	Green	Green	Green	Amber	Green	Green

Source: CEPA analysis

Note: The reported VIF excludes the squared density term

***Key**: Green Amber Red*

Table 6.3: Phase 2 assessment results for the WRP2 models

Driver	WRP2 Model 1	WRP2 Model 3	WRP2 Model 4	WRP2 Model 5	WRP2 Model 7	WRP2 Model 8	WRP2 Model 24	WRP2 Model 23
	PR19	MSOA instead of LAD density	LAD from MSOA density	Prop. per length instead of LAD density	Alternative weights 1 (1,2,3,3,4,4,5)	Alternative weights 2 (1,4,9,9,16,16,25)	% DI from pumped reservoirs	% DI from reservoirs
Properties (log)	1.069***	1.057***	1.075***	1.027***	1.075***	1.079***	1.074***	1.032***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Weighted average complexity (log)	0.377	0.315	0.343	0.365			0.264	0.258
	{0.123}	{0.234}	{0.183}	{0.143}			{0.297}	{0.324}
Density variable (log) [35]	-1.412***	-5.048**	-1.468**	-7.440**	-1.402***	-1.420***	-0.743	-1.122**
	{0.005}	{0.034}	{0.026}	{0.030}	{0.007}	{0.008}	{0.154}	{0.021}
Squared density variable (log)	0.087***	0.306**	0.091**	0.810**	0.086**	0.087**	0.037	0.068**
	{0.009}	{0.033}	{0.031}	{0.042}	{0.011}	{0.014}	{0.296}	{0.033}
Alternative WAC 1 (log)					0.430**			
					{0.011}			
Alternative WAC 2 (log)						0.258**		
						{0.035}		
DI from pumped reservoirs (%)							0.006***	
							{0.005}	
DI from reservoirs (%)								0.006***
								{0.001}
Constant	-5.805***	9.591	-5.660***	6.136	-5.882***	-5.990***	-8.033***	-6.392***
	{0.000}	{0.286}	{0.002}	{0.389}	{0.000}	{0.000}	{0.000}	{0.000}

[35] The models use the PR19 LAD density variable unless otherwise specified in the column header.

Driver	WRP2 Model 1	WRP2 Model 3	WRP2 Model 4	WRP2 Model 5	WRP2 Model 7	WRP2 Model 8	WRP2 Model 24	WRP2 Model 23
R-squared	0.907	0.896	0.902	0.905	0.906	0.904	0.896	0.900
RESET test	0.324	0.729	0.367	0.203	0.381	0.389	0.545	0.594
VIF (max)*	1.220	1.308	1.253	1.158	1.176	1.156	1.434	1.359
Pooling	0.999	1.000	0.999	0.997	0.998	0.999	0.99	0.997
Normality	0.574	0.416	0.812	0.527	0.733	0.794	0.679	0.445
Heteroskedasticity	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Minimum efficiency score	0.500	0.473	0.501	0.484	0.51	0.508	0.476	0.463
Maximum efficiency score	1.979	1.983	1.986	1.948	1.983	1.984	2.193	2.343
Sensitivity of estimated coefficients to removal of most and least efficient company	Green	Amber	Green	Green	Green	Green	Green	Green
Sensitivity of estimated coefficients to removal of first and last year of sample	Green	Green	Green	Green	Green	Green	Green	Green

Source: CEPA analysis

Note: The reported VIF excludes the squared density term

***Key**:* *Green* *Amber* *Red*

In summary, our statistical tests show that:

- All estimated coefficients are significant and have the expected sign, with the exception of the PR19 weighted average complexity driver, which is not significant at the 10% level (p value = 0.123) in the PR19 model specification or in any of the alternative model specifications tested.
- All the alternative model specifications have an explanatory power (R^2) that is lower than the PR19 model specifications.
- The models all perform well on the RESET, VIF, pooling and normality tests. The LM test for all models is supportive of a Random Effects specification.
- All models fail the heteroskedasticity tests. We do not consider the results of these tests to be problematic, due to the use of robust standard errors.
- We find that all the models are robust to the removal of companies and years from the sample with the exception of the WRP1 model including separate surface water and ground water treatment complexity bands, as discussed in more detail below.

We set out our overall conclusions of our Phase 2 assessment of WRP models below.

For the **alternative density variables**, we find that there are no improvements in the statistical performance of the WRP models using the alternative density variables:

- The explanatory power of the models is slightly lower than that of the PR19 models and the statistical significance of the coefficients for the alternative density variables is also often slightly lower than that of the PR19 density variables. This is particularly so for the MSOA density variable.
- The MSOA density driver loses some significance (going from significant at the 5% level to significance at the 10% level) with the removal of SSC (as the most efficient company) in the WRP2 model.
- The introduction of the MSOA density driver produces swings in efficiency scores up to 45pp compared to the PR19 model for one company (WSX). Seven other companies show a change in efficiency score exceeding 5pp, with a maximum being 16pp. The introduction of the properties per length of mains density driver produces changes in efficiency scores of up to 27pp compared to the PR19 model specification, with six companies showing a change of over 5pp.

However, we understand that Ofwat's rationale for developing alternative density variables based on the MSOA dataset, is that this uses more granular data that may provide a more accurate picture of the relative density between company areas and may be less sensitive to changes in the dataset over time. These density variables also do not rely on companies' mapping of LADs to company areas, which was the method used for the PR19 density variable. This can improve the transparency and reliability of the dataset. Additionally, similar to the MSOA density variables, the property per length of mains variable is statistically significant and performs similar as the PR19 LAD measure in terms of statistical robustness of the model. However, this variable is, to some extent, more endogenous compared to the other density variables, as the length of mains is partially under companies' control.

We consider that the alternative MSOA density variables could be considered an improvement over the PR19 specification provided the underlying dataset for the calculation of the density variables is robust and stable and they continue to produce statistically robust results when additional outturn data is added. Due to potential issues with endogeneity, we also consider that the alternative MSOA density variables are better than the property per length of mains density variable. However, if the underlying MSOA dataset appears to be insufficiently robust and stable, property per length of mains could be an alternative to overcome the problems associated with the PR19 LAD measure.

We find that the inclusion of **separate surface water and ground water treatment complexity bands** in the WRP1 model performs relatively well in terms of the statistical significance of the coefficients and performance against statistical tests. However, the explanatory power of the WRP1 model is slightly lower than that of the PR19 models. Furthermore, the results are sensitive to removing years from the sample. The surface water variable becomes insignificant when the first year of the sample is excluded.

In addition, given that these variables use the same bands as the PR19 driver, it is not clear that there is sufficient technical justification or benefit from using two variables instead of one. As we set out in our Phase 1 assessment, the coefficient on surface water bands is only statistically significant and has an intuitive sign for bands 3-6, but not for higher bands. We find this surprising given that almost all surface water is treated in bands 3-6, which would suggest that this variable is not a good candidate for explaining variation in costs between companies or over time. Therefore, we do not consider that this option leads to a clear improvement compared over the PR19 model specification.

For the **alternative weighted average complexity variable** in WRP2:

- We find that the PR19 weighted average complexity variable in WRP2 is not significant when using the latest dataset. When excluding the last year of data (2021-22), the PR19 weighted average complexity variable in WRP2 is significant at the 10% level[36]. It is also significant (at the 5% level) when we exclude the most efficient company (SSC). However, the variable is still insignificant at the 10% level when we exclude the least efficient company (SRN) or the first year of data.
- While the alternative WAC variables 1 and 2 perform better than the PR19 variable in terms of the statistical significance of the estimated coefficients, adopting these variables in the PR24 models would require further evidence that the proposed weights are justified based on the relative costs of treating water in different complexity bands. The PR19 WAC variable assumes that there is a linear relationship between the cost of treating water at different complexity levels. The alternative WAC variable proposed by Severn Trent Water assumes that there are similar costs associated with treating water at complexity levels 2 and 3, as well as levels 4 and 5. Based on our discussions with engineers, we have not been able to confirm that Severn Trent's assumptions apply widely across the industry. Therefore, Ofwat would need to consider if there is sufficient evidence to justify the use of alternative weights before amending the approach used at PR19.
- We also note that our analysis indicates that the use of these alternative weights does not have a major impact on efficiency scores compared to the PR19 model specification.

For the **water resource drivers,** we find that there are no improvements in the statistical performance of the WRP models nor there is sufficient technical justification for including these variables in the PR24 models:

- The "% of DI from pumped reservoirs" variable is statistically significant in both WRP models, performs well against the statistical and sensitivity tests we applied and has a strong technical rationale as pumped storage reservoirs are expected to be the most expensive water sources to operate (e.g., due to pumping) and maintain (e.g., due to reservoir maintenance requirements under the reservoir act).
- However, the PR19 density drivers in both WRP models lose significance when including % of DI from pumped reservoirs. While the overall correlation between these variables is not very strong, Thames Water (TMS) (the densest network) has the largest % of DI from pumped reservoirs, while Wessex Water (WSX) (one of the sparsest networks) has the lowest % DI from pumped reservoirs. We tested a model with % of DI from pumped reservoirs as an explanatory variable but excluding TMS from the sample, to check if the results might be driven by TMS as an outlier. We found similar results, with the % DI from pumped reservoirs still significant while both density variables are no longer significant. Overall, we do not consider that the inclusion of % DI from pumped reservoirs improves the performance of the models given its interaction with the density terms and the observed decrease in the R^2.
- The "% of DI from reservoirs" variable is also statistically significant in both WRP models and does not result in the density variables losing significance. However, we do not consider that the inclusion of % DI from reservoirs

[36] P-value = 0.067

is sufficiently justified from an engineering perspective. Reservoirs may have higher maintenance costs than other sources because of requirements under the reservoir act. However, engineering rationale suggests that number of reservoirs would be a better driver of these costs than % of DI from reservoirs. Impounding reservoirs (which are a subset of total reservoirs) also require less pumping to water treatment works than other sources, so it is not clear that the expected relationship between "% of DI from reservoirs" and water resources plus base costs is positive. There is also a risk that "% DI from reservoirs" is picking up a treatment complexity effect, which is reducing the statistical significance of the weighted average complexity driver (see Models 25 and 26).

- We also note that the introduction of % DI from reservoirs and, to a slightly lesser extent, of % DI from pumped reservoirs, widens the range of efficiency scores largely due to an increase in the maximum efficiency score estimated by the models. The introduction of % DI from pumped reservoirs produces changes in efficiency scores of up to 24pp, and for 13 of the 17 water companies' efficiency scores change by more than 5pp. The introduction of % DI from reservoirs produces changes in efficiency scores of up to 36pp, and 12 of the 17 water companies' efficiency scores change by more than 5pp.

6.1.2. Treated water distribution models

We present the full statistical results of our Phase 2 assessment for the treated water distribution models in Table 6.4.

Table 6.4: Phase 2 assessment results for the TWD model

Driver	TWD Model 1	TWD Model 2	TWD Model 3	TWD Model 4	TWD Model 5	TWD Model 6	TWD Model 17
	PR19	APH (dist), no booster stations	APH (dist), with booster stations	MSOA instead of LAD density	LAD from MSOA density	Prop. per length instead of LAD density	Time trend
Lengths of main (log)	1.077***	1.069***	1.069***	1.026***	1.070***	1.072***	1.070***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Boosters per length (log)	0.437***		0.333***	0.433***	0.461***	0.488***	0.409***
	{0.002}		{0.008}	{0.001}	{0.002}	{0.001}	{0.006}
Density term (log)[37]	-2.946***	-3.203***	-2.879***	-5.561***	-2.729***	-14.921***	-2.739***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Squared density term (log)	0.235***	0.245***	0.228***	0.393***	0.219***	1.898***	0.218***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Average Pumping Head – distribution (log)		0.313***	0.276***				
		{0.000}	{0.000}				
Year							0.011***
							{0.009}
Constant	4.723***	2.892*	3.024**	15.638***	4.155***	25.065***	-18.487**
	{0.002}	{0.057}	{0.032}	{0.002}	{0.008}	{0.000}	{0.028}
R-squared	0.957	0.960	0.964	0.952	0.955	0.958	0.958
RESET test	0.102	0.599	0.476	0.122	0.090	0.489	0.315
VIF (max)	2.108	1.026	2.163	1.592	1.833	1.864	2.117
Pooling	0.813	0.824	0.794	0.873	0.799	0.903	0.991

[37] The models use the PR19 LAD density variable unless otherwise specified in the column header.

Driver	TWD Model 1	TWD Model 2	TWD Model 3	TWD Model 4	TWD Model 5	TWD Model 6	TWD Model 17
Normality	0.520	0.918	0.926	0.014	0.072	0.738	0.470
Heteroskedasticity	0.246	0.474	0.883	0.046	0.132	0.004	0.123
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Minimum efficiency score	0.771	0.714	0.732	0.750	0.795	0.737	0.738
Maximum efficiency score	1.376	1.330	1.358	1.425	1.400	1.378	1.306
Sensitivity of estimated coefficients to removal of most and least efficient company	+	+	+	+	+	+	+
Sensitivity of estimated coefficients to removal of first and last year of sample	+	+	+	+	+	+	+

Source: CEPA analysis

Note: The reported VIF excludes the squared density term

***Key**:* + *Green* – *Amber* × *Red*

In summary, our statistical tests show that:

- All estimated coefficients are significant and have the expected sign.
- All models perform well on the RESET, pooling and VIF test. The LM test for all models is supportive of a Random Effects specification.
- The model using MSOA as the density driver fails the normality and heteroskedasticity tests, and the model using properties per length of mains as the density driver fails the heteroskedasticity test. However, we do not consider the failure of the normality or the heteroskedasticity test to be problematic, due to the use of a Random Effects modelling approach and robust standard errors, respectively.
- We find that all the models are robust to the removal of companies and years from the sample.

We set out our overall conclusions of our Phase 2 assessment of TWD models below.

We find that the inclusion of **APH** in the TWD model results in a small improvement in the overall performance compared to the PR19 model specification:

- The models including APH are statistically robust and they result in small improvements in explanatory power compared to the PR19 model specifications. The inclusion of APH is also supported by a strong technical rationale as a measure of the level of pumping required, and thus the amount of energy used by water companies in pumping water around their networks.
- We note that, including APH, both instead of and alongside booster stations per length, results in large changes in efficiency scores for individual companies. Individual companies show swings in efficiency scores up to -23pp and +18pp compared to the PR19 TWD model. PRT shows the largest increase in efficiency score (+18pp), while SES shows the largest decrease in efficiency score (-23pp). This probably suggests that APH is capturing the hilliness of the region better than BPL (e.g., PRT operates in a relatively flat area and reports low levels APH, while SES reports high levels of APH compared to the industry average).
- While the models including APH perform well in terms of statistical robustness and engineering rationale, one potential concern is around the quality of the APH data. Ofwat considers that the TWD APH data is of better quality than the other APH data based on findings from the Turner & Townsend study.[38] While the quality of the APH data has improved in the last few years, Ofwat may wish to consider if the APH data is sufficiently robust to use in the PR24 models once an additional year of data becomes available. If this criterion is satisfied, we consider the inclusion of this variable is justified based on our selection criteria.

We also note that the use of **booster stations per length alongside APH** produces robust statistical results and improves the explanatory power of the models more so than the use of APH without boosters per length. As both APH and boosters per length aim to control for network topography, this finding is somewhat surprising, and the engineering rationale for these results should be explored further. As APH is likely to control for energy costs, it is possible that the inclusion of a booster stations per length variable is more likely to provide a measure of the asset intensity of a network.

The decision on whether to include boosters per length alongside APH can have a material impact on the estimated efficiency scores for the TWD model. The largest changes in efficiency scores (up to 23pp) occur in the model without boosters per length which suggests that these changes are driven primarily by the exclusion of boosters, rather than the inclusion of APH. In the model with APH and no booster stations per length, 13 of the 17 water companies experience changes in efficiency scores greater than 5pp compared to the PR19 models. In contrast, when booster stations are included alongside APH, only eight companies see efficiency scores change by more

[38] Turner & Townsend, WRC (24 March 2022), Average Pumping Head: data quality improvement.

URL: https://www.ofwat.gov.uk/wp-content/uploads/2022/05/Average-Pumping-Head-Data-Quality-Improvement-Final-Report-.pdf than 5pp. Ofwat could therefore consider a mix of models (i.e., some that include APH, and some that include booster stations per length of main). This may lead to a similar outcome to when booster stations per length of main and APH are included in the same model.

As was the case for the WRP models, we find that there are no improvements in the statistical performance of the TWD models using the **alternative density variables**. All alternative density models produce results which, from a statistical perspective, are similar to the PR19 specification:

- We note that the introduction of MSOA as the density driver produces changes in efficiency scores of up to 19pp compared to the PR19 model specification, with ten companies exhibiting changes in excess of 5pp. The overall range of efficiency scores in the model using MSOA density also increases by 7pp relative to the PR19 model specification.
- The introduction of LAD from MSOA results in more muted changes in efficiency scores compared to the PR19 density measure. This is likely due to the fact that the MSOA density aggregated to LAD level gives a view of relative densities which is more similar to the PR19 measure than the disaggregated MSOA measure. Only two companies display changes in efficiency scores larger than 5pp compared to the PR19 model specification, with the largest change equal to 7pp.
- The introduction of properties per length of mains as the density driver produces swings in efficiency scores of up to 17pp compared to the PR19 model specification, with six companies showing changes larger than 5pp.

As noted previously, we understand that Ofwat's rationale for developing alternative density variables based on the MSOA dataset, is that this uses more granular data that may provide a more accurate picture of the relative density between company areas and may be less sensitive to changes in the dataset over time. These density variables also do not rely on companies' mapping of LADs to company areas, which was the method used for the PR19 density variable. This can improve the transparency and reliability of the dataset.

Additionally, similar to the MSOA density variables, the property per length of mains variable is statistically significant and performs similar as the PR19 LAD measure in terms of robustness. However, this variable is, to some extent, more endogenous compared to the other density variables, as the length of main is partially under companies' control.

We consider that the alternative MSOA density variables might be an improvement over the PR19 specification and form the new baseline, provided the underlying dataset for the calculation of the density variables is robust and stable and continues to produce statistically robust results when additional outturn data is added. Due to potential issues with endogeneity, we also consider that the alternative MSOA density variables are better than the property per length of mains density variable. However, if the underlying MSOA dataset appears to be insufficiently robust and stable, property per length of mains could be an alternative to overcome the problems associated with the PR19 LAD measure.

The inclusion of a linear **time trend** is statistically significant in TWD model but does not result in a clear overall improvement in the statistical performance of the model (i.e., no material change in R-squared or significance of coefficients and all robustness tests are passed with or without the time trend). As the time trend coefficient is positive, including a time trend in the regression would allow costs to gradually increase over time. The impact of this is reflected in the fact that the inclusion of a linear time trend in the TWD model improves efficiency scores for 16 out of the 17 water companies. This is expected because the positive coefficient on the time trend variable provides an uplift to modelled costs. The changes in efficiency scores compared to the PR19 model specification are between 3pp and 7pp with two companies seeing changes greater than 5pp. It is important to note that the overall impact on allowances is ambiguous, as the inclusion of a time trend is likely to result in a higher catch-up efficiency challenge.

A potential rationale for including a time trend could be to reflect increasing sector costs over time that are not explained by the other explanatory variables. For instance, increasing leakage costs has been mentioned by some companies. However, as the time trend is a catch-all variable it is not possible to say exactly what is driving the estimated effect. This means that careful consideration is needed to decide whether the rationale is sufficiently strong from an engineering perspective, and if the increasing costs will continue into the future.

6.1.3. Wholesale water models

We present the full results of our Phase 2 assessment for the wholesale water models in Table 6.5 and Table 6.6.

Table 6.5: Phase 2 assessment results for the WW1 model

Driver	WW1 Model 1	WW1 Model 2	WW1 Model 2 (a)	WW1 Model 4	WW1 Model 4 (a)	WW1 Model 5	WW1 Model 6	WW1 Model 7	WW1 Model 12	WW1 Model 16	WW1 Model 23	WW1 Model 22
	PR19	APH (all), no booster stations	APH (TWD), no booster stations	APH (all), with booster stations	APH (TWD), with booster stations	MSOA instead of LAD density	LAD from MSOA density	Prop. Per length instead of LAD density	Separate SW / GW	Length of mains instead of prop.	% DI from pumped reservoirs	% DI from reservoirs
Properties (log)	1.071*** {0.000}	1.096*** {0.000}	1.066*** {0.000}	1.093*** {0.000}	1.067*** {0.000}	1.052*** {0.000}	1.072*** {0.000}	1.044*** {0.000}	1.092*** {0.000}		1.072*** {0.000}	1.051*** {0.000}
Water treated in bands 3-6 (%)	0.004*** {0.000}	0.003* {0.059}	0.004*** {0.005}	0.003** {0.032}	0.004*** {0.002}	0.003** {0.011}	0.003*** {0.002}	0.003*** {0.001}		0.004*** {0.000}	0.003** {0.017}	0.003*** {0.007}
Boosters per length (log)	0.335** {0.032}			0.318** {0.017}	0.249* {0.055}	0.509*** {0.003}	0.457*** {0.008}	0.377** {0.033}	0.223* {0.075}	0.244* {0.081}	0.448** {0.021}	0.390** {0.038}
Density term (log)[39]	-2.094*** {0.000}	-2.579*** {0.000}	-2.321*** {0.000}	-2.250*** {0.000}	-2.075*** {0.000}	-4.684*** {0.001}	-1.849*** {0.000}	-11.26*** {0.000}	-2.201*** {0.000}	-2.446*** {0.000}	-1.515*** {0.000}	-1.826*** {0.000}
Squared density term (log)	0.147*** {0.000}	0.177*** {0.000}	0.157*** {0.000}	0.159*** {0.000}	0.144*** {0.000}	0.301*** {0.000}	0.132*** {0.000}	1.318*** {0.000}	0.152*** {0.000}	0.192*** {0.000}	0.106*** {0.000}	0.129*** {0.000}
Average Pumping Head (log)		0.323*** {0.008}	0.279*** {0.004}	0.297*** {0.007}	0.243** {0.012}							
Surface water treated in bands 3-6 (%)									0.006*** {0.000}			
Groundwater treated in bands 3-6 (%)									0.004*** {0.000}			
Lengths of main (log)										1.052*** {0.000}		
DI from pumped reservoirs (%)											0.004***	

[39] The models use the PR19 LAD density variable unless otherwise specified in the column header.

Driver	WW1 Model 1	WW1 Model 2	WW1 Model 2 (a)	WW1 Model 4	WW1 Model 4 (a)	WW1 Model 5	WW1 Model 6	WW1 Model 7	WW1 Model 12	WW1 Model 16	WW1 Model 23	WW1 Model 22
											{0.002}	
DI from reservoirs (%)												0.003*** {0.002}
Constant	-1.565* {0.074}	-2.888* {0.072}	-3.028*** {0.010}	-2.838*** {0.005}	0.243** {0.012}	10.300* {0.056}	-1.958 {0.206}	15.655*** {0.003}	-2.300*** {0.002}	3.058** {0.031}	-3.140*** {0.003}	-2.085** {0.014}
R-squared	0.970	0.969	0.971	0.972	0.973	0.963	0.965	0.965	0.972	0.969	0.970	0.968
RESET test	0.223	0.786	0.827	0.597	0.658	0.178	0.164	0.205	0.346	0.152	0.157	0.257
VIF (max)	2.211	1.506	2.214	2.296	2.271	1.789	1.955	1.879	2.282	2.115	2.212	2.239
Pooling	0.869	0.592	0.758	0.854	0.923	0.987	0.940	0.962	0.88	0.702	0.809	0.907
Normality	0.225	0.069	0.032	0.178	0.461	0.510	0.268	0.445	0.071	0.359	0.241	0.223
Heteroskedasticity	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Minimum efficiency score	0.777	0.809	0.778	0.814	0.791	0.735	0.764	0.714	0.752	0.788	0.832	0.814
Maximum efficiency score	1.385	1.333	1.481	1.300	1.373	1.527	1.490	1.411	1.303	1.374	1.500	1.466
Sensitivity of estimated coefficients to removal of most and least efficient company	Green	Green	Amber	Green	Amber	Amber	Green	Amber	Green	Green	Green	Green
Sensitivity of estimated coefficients to removal of first and last year of sample	Green	Green	Green	Green	Green	Green	Green	Green	Green	Amber	Green	Green

Source: CEPA analysis

Note: The reported VIF excludes the squared density term

***Key**:* *Green* *Amber* *Red*

Table 6.6: Phase 2 assessment results for the WW2 model

Driver	WW2 Model 1	WW2 Model 2	WW2 Model 2 (a)[40]	WW2 Model 4	WW2 Model 4 (a)[41]	WW2 Model 5	WW2 Model 6	WW2 Model 7	WW2 Model 9	WW2 Model 10	WW2 Model 11
	PR19	**APH (all), no booster stations**	**APH (TWD), no booster stations**	**APH (all), with booster stations**	**APH (TWD), with booster stations**	**MSOA instead of LAD density**	**LAD from MSOA density**	**Prop. Per length instead of LAD density**	**Alternative weights 3 (1,4,9,16,25,36,49)**	**Alternative weights 2 (1,4,9,9,16,16,25)**	**Alternative weights 3 (1,4,9,16,25,36,49)**
Properties (log)	1.059*** {0.000}	1.087*** {0.000}	1.057*** {0.000}	1.083*** {0.000}	1.057*** {0.000}	1.046*** {0.000}	1.061*** {0.000}	1.036*** {0.000}	1.068*** {0.000}	1.071*** {0.000}	1.059*** {0.000}
WAC (log)	0.430*** {0.001}	0.280* {0.094}	0.371** {0.021}	0.309* {0.052}	0.394*** {0.008}	0.322** {0.030}	0.354** {0.016}	0.366*** {0.007}			
Boosters per length (log)	0.334** {0.019}			0.319** {0.012}	0.259** {0.038}	0.486*** {0.003}	0.444*** {0.005}	0.351** {0.033}	0.357** {0.017}	0.357** {0.021}	0.329** {0.022}
Density term (log) [42]	-1.832*** {0.000}	-2.401*** {0.000}	-2.104*** {0.000}	-2.051*** {0.000}	-1.854*** {0.000}	-4.308*** {0.002}	-1.648*** {0.001}	-10.322*** {0.000}	-1.838*** {0.000}	-1.843*** {0.000}	-1.812*** {0.000}
Squared density term (log)	0.128*** {0.000}	0.164*** {0.000}	0.142*** {0.000}	0.145*** {0.000}	0.128*** {0.000}	0.276*** {0.001}	0.117*** {0.000}	1.201*** {0.000}	0.129*** {0.000}	0.129*** {0.000}	0.126*** {0.000}
APH (log)		0.309*** {0.008}	0.270*** {0.009}	0.275** {0.012}	0.227** {0.035}						
Alternative WAC 1 (log)									0.422*** {0.001}		
Alternative WAC 2 (log)										0.251*** {0.008}	
Alternative WAC 3 (log)											0.283*** {0.002}

[40] As there are some concerns regarding the quality of APH (WRP) data, and this driver was not statistically significant in the WRP models, we have also included APH (TWD) instead of APH (all) as a driver in the WW models.

[41] As there are some concerns regarding the quality of APH (WRP) data, and this driver was not statistically significant in the WRP models, we have also included APH (TWD) instead of APH (all) as a driver in the WW models.

[42] The models use the PR19 LAD density variable unless otherwise specified in the column header.

Driver	WW2 Model 1	WW2 Model 2	WW2 Model 2 (a)[40]	WW2 Model 4	WW2 Model 4 (a)[41]	WW2 Model 5	WW2 Model 6	WW2 Model 7	WW2 Model 9	WW2 Model 10	WW2 Model 11
Lengths of main (log)											
DI from pumped reservoirs (%)											
DI from reservoirs (%)											
Constant	-2.589*** {0.001}	-3.486** {0.021}	-3.833*** {0.001}	-3.476*** {0.000}	-3.693*** {0.000}	8.674 {0.108}	-2.795* {0.064}	13.516*** {0.008}	-2.497** {0.010}	-2.635** {0.011}	-2.904*** {0.001}
R-squared	0.971	0.969	0.970	0.973	0.973	0.965	0.967	0.968	0.968	0.967	0.970
RESET test	0.122	0.824	0.771	0.543	0.973	0.075	0.075	0.072	0.134	0.128	0.100
VIF (max)	2.214	1.584	1.232	2.315	2.216	1.741	1.948	1.868	2.209	2.214	2.225
Pooling	0.724	0.534	0.701	0.749	0.877	0.965	0.862	0.958	0.704	0.753	0.757
Normality	0.838	0.439	0.441	0.607	0.928	0.574	0.583	0.483	0.642	0.743	0.951
Heteroskedasticity	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Minimum efficiency score	0.784	0.766	0.744	0.788	0.7558	0.737	0.758	0.718	0.752	0.741	0.756
Maximum efficiency score	1.414	1.317	1.450	1.327	1.337	1.532	1.500	1.422	1.423	1.438	1.432
Sensitivity of estimated coefficients to removal of most and least efficient company	+	−	+	+	−	−	+	−	+	+	+
Sensitivity of estimated coefficients to removal of first and last year of sample	+	−	+	+	+	+	+	+	+	+	+

Driver	WW2 Model 12	WW2 Model 19	WW2 Model 18
	Length of mains instead of prop.	% DI from pumped reservoirs	% DI from reservoirs
Properties (log)		1.063*** {0.000}	1.044*** {0.000}
WAC (log)	0.383*** {0.006}	0.316** {0.032}	0.332** {0.019}
Boosters per length (log)	0.244* {0.053}	0.429** {0.019}	0.377** {0.024}
Density term (log) [43]	-2.222*** {0.000}	-1.383*** {0.000}	-1.647*** {0.000}
Squared density term (log)	0.175*** {0.000}	0.097*** {0.000}	0.116*** {0.000}
APH – all (log)			
Alternative WAC 1 (log)			
Alternative WAC 2 (log)			
Alternative WAC 3 (log)			
Lengths of main (log)	1.042*** {0.000}		
DI from pumped reservoirs (%)		0.003** {0.012}	
DI from reservoirs (%)			0.002** {0.011}
Constant	2.127 {0.113}	-3.730*** {0.000}	-2.856*** {0.000}
R-squared	0.970	0.970	0.969
RESET test	0.084	0.141	0.213

[43] The models use the PR19 LAD density variable unless otherwise specified in the column header.

Driver	WW2 Model 12	WW2 Model 19	WW2 Model 18
VIF (max)	2.126	2.217	2.251
Pooling	0.573	0.582	0.768
Normality	0.448	0.800	0.775
Heteroskedasticity	0.000	0.000	0.000
Pooled OLS vs RE (LM test)	0.000	0.000	0.000
Minimum efficiency score	0.782	0.812	0.799
Maximum efficiency score	1.346	1.508	1.477
Sensitivity of estimated coefficients to removal of most and least efficient company	Green	Green	Green
Sensitivity of estimated coefficients to removal of first and last year of sample	Green	Green	Green

Source: CEPA analysis

Note: The reported VIF excludes the squared density term

***Key**: Green Amber Red*

In summary, our statistical tests show that:

- All estimated coefficients are significant and have the expected sign.
- There are very marginal improvements in the explanatory power of some of the alternative model specifications. For example, the inclusion of APH (all) alongside booster stations per length improves the R^2 by 0.002 for both WW1 and WW2.
- All models pass the RESET, pooling, VIF and normality tests. The LM test for all models is supportive of a Random Effects specification.
- All models fail the heteroskedasticity test, however, we do not consider this test to be problematic, due to the use of robust standard errors.
- We find that all the models are robust to the removal of companies and years from the sample with the exception of the WW2 including APH (all) without boosters per length, WW2 including APH (TWD) with boosters per length and some of the models including MSOA density, where some of the variables in the models lose statistical significance when observations are excluded from the sample. These instances are discussed in more detail below.

We set out our overall conclusions of our Phase 2 assessment of WW models below, building on our assessment of WRP and TWD models.

Unlike the TWD model, the inclusion of **APH** (all or TWD) in the WW models does not produce improvements in overall performance compared to the PR19 model specification:

- The APH models have similar explanatory power to the PR19 models and tend to cause the water treatment complexity driver to lose significance (particularly when APH (all) is used). However, the inclusion of APH is supported by a strong technical rationale.
- The inclusion of APH (all) in the WW1 model causes the water treatment complexity driver (bands 3-6) to lose some statistical significance, especially when the booster stations per length variable is excluded (the water treatment variable becomes significant only at the 10% level in this case). Similarly, the inclusion of APH causes the WAC variable in the WW2 model to lose some significance (significant at the 10% level), both when APH is used alongside and instead of booster stations. In the WW2 model using APH (all) without boosters per length, the water treatment variable is insignificant when the first year is excluded, and when the most efficient company (SSC) is excluded.
- The WW models using APH (TWD) seem to perform better in our sensitivity analysis compared to the models including APH (all). Unlike the model with APH (all), the water treatment complexity driver (bands 3-6) does not lose significance when including APH (TWD) compared to the PR19 model. The WAC variable and the water treatment variable also do not lose significance when removing individual years of data or the least/most efficient company. However, boosters per length becomes insignificant when excluding the least efficient company in the WW1 and WW2 models including APH (TWD).
- As seen in the TWD models, the inclusion of APH (TWD and all) has a large impact on efficiency scores. The introduction of APH (all) produces changes in efficiency scores up to 21pp compared to the PR19 model, and 12 out of 17 water companies show swings exceeding 5pp. The introduction of APH (TWD) produces similar swings, with changes in efficiency scores up to 22pp compared to the PR19 model. However, only 7 out of 17 water companies show swings exceeding 5pp.

We consider that APH works better in the TWD model compared to the WW models both in terms of statistical performance and engineering rationale. This can be explained by the fact that most pumping costs are related to treated water distribution. In the WW models, we consider that APH (TWD) performs marginally better compared to APH (all). This is based not only on statistical performance but also on the fact that we understand that the quality of APH (TWD) data is somewhat better than APH (WRP) and our results showed that APH (WRP) was insignificant in the WRP models. Therefore, we expect APH (TWD) to be a more robust measure of APH than APH (all). Overall, we consider that there is a strong rationale for including APH in at least one of the WW models, if Ofwat decides to include APH in the TWD model as the main variable capturing network topography.

The use of **booster stations per length alongside APH** also produces robust statistical results. As explained in the discussion regarding TWD models, the engineering rationale for these results should be explored further. We note that the decision on whether to include boosters per length alongside APH can have a material impact on the estimated efficiency scores for the WW models. Ofwat could therefore consider a mix of models (i.e., some that include APH, and some that include booster stations per length of main). This may lead to a similar outcome to when booster stations per length of main and APH are included in the same model.

In line with the findings for the WRP and TWD models, we find that there are no improvements in the statistical performance of the WW models using the **alternative density variables.** The explanatory power of the models with alternative density variables is slightly lower than that of the PR19 models, particularly in the case of the MSOA density variable. In addition, in the models using MSOA as the density driver, the water treatment variables are insignificant when PRT (most efficient) and SES (least efficient) companies are excluded.

As noted previously, we understand that Ofwat's rationale for developing alternative density variables based on the MSOA dataset, is that this uses more granular data that may provide a more accurate picture of the relative density between company areas and may be less sensitive to changes in the dataset over time. These density variables also do not rely on companies' mapping of LADs to company areas, which was the method used for the PR19 density variable. This can improve the transparency and reliability of the dataset.

Additionally, similar to the MSOA density variables, the property per length of mains variable is statistically significant and performs similar as the PR19 LAD measure in terms of robustness. However, this variable is, to some extent, more endogenous compared to the other density variables, as the length of main is partially under companies' control.

We consider that the alternative MSOA density variables might be an improvement over the PR19 specification and form the new baseline, provided the underlying dataset for the calculation of the density variables is robust and stable and they continue to produce statistically robust results when additional outturn data is added. Due to potential issues with endogeneity, we also consider that the alternative MSOA density variables are better than the property per length of mains density variable. However, if the underlying MSOA dataset appears to be insufficiently robust and stable, property per length of mains could be an alternative to overcome the problems associated with the PR19 LAD measure.

For the WRP models, we concluded that there is insufficient justification for the inclusion of **separate surface water and ground water treatment complexity bands** despite relatively good statistical performance. In the WW models, we also find that these variables perform relatively well from a statistical perspective, however, our overall assessment of this option does not change.

For the WRP models, we concluded that the inclusion of **% DI from pumped reservoirs** or **% DI from reservoirs** does not sufficiently improve the performance of the models and/or is insufficiently supported by a technical rationale. In the WW models, we find that the inclusion of % DI from pumped reservoirs does not result in a loss of significance for the density variables, as was the case in the WRP models. However, the fact that the inclusion of this variable, which is specifically targeted at water resource costs, does not improve the overall performance of the WRP models means that it should not be considered a viable alternative option in the WW models. In the case of **% DI from reservoirs**, we apply the same assessment as in the WRP section where we concluded that the technical rationale and the expected relationship between prevalence of reservoirs and costs is too ambiguous,

We find that the inclusion of **length of mains** as a scale driver in the WW models instead of number of properties produces robust statistical results but does not offer a clear improvement over the PR19 model specification.

The use of length of mains has some implications for the use of booster per length in the WW models:

- The boosters per length variable becomes statistically significant only at the 10% level when length of mains is used as the scale driver in both WW1 and WW2 models.

- In the WW1 model using lengths of mains as the scale driver, boosters per length becomes insignificant when removing the first year.

Overall, we consider that Ofwat could consider this as a possible option for the PR24 models given that length of mains is the scale driver for the TWD model and TWD costs make up the largest proportion of WW costs. It is also worth noting that using lengths of mains as the scale driver produces changes in efficiency scores up to 20pp, and 10 out of 17 water companies show swings exceeding 5pp.

6.2. Wastewater models

Table 6.7 summarises the results of our Phase 2 assessment, which considers the variables we progressed from our Phase 1 assessment, whose findings are summarised in Table 5.2.

For the sewage collection and sewage treatment models, the PR19 models are used as the baseline for comparison. For the wholesale wastewater network plus models, the baseline model uses load, pumping capacity and ammonia consents as the core cost drivers, as the best performing model from our exploratory modelling as presented in detail in Section 5.2.

The purpose of the Phase 2 assessment is to identify options that could improve the performance of Ofwat's base cost models relative to the PR19 costs modelling suite. For this reason, for the two sewage collection and sewage treatment models in the existing PR19 modelling suite, we only recommend that Ofwat considers a model if it provides a demonstrable improvement on the PR19 baseline. However, for the development of the new Wholesale Wastewater Network Plus model which has no PR19 modelling baseline, we recommend that Ofwat considers any models that perform as well as the baseline model specification established in our Phase 1 assessment (WWWN+ model 21).

For the more aggregated models (WWWN+), a suitable driver for consideration will be one that also performs well in the more disaggregated model.

Table 6.7: Variables considered in our Phase 2 assessment

Model	SWC1	SWC2	SWT1	SWT2	WWWN+
Rainfall					
Urban rainfall per sewer length (log)	−	−			−
Treatment complexity					
BOD consents ≤ 7mg/l (%)			✕	✕	
Network topography and coastal population					
'Coastal' population (%)			✕	✕	
Pumping capacity per sewer length			✕	✕	
Sewage treatment economies of scale					
STW size cut-offs				+	+
Weighted Average Band (Anglian Water)				−	✕
Weighted Average Band (CEPA)				+	+
Weighted average treatment size				+	−
Density					
MSOA density, aggregated at LAD level		−			
MSOA density		−			

Key: + *Consider* − *Needs further consideration* ✕ *Do not consider*

In the following subsections, we present the results of our Phase 2 assessment by Botex category.

6.2.1. Sewage collection models

We present the full results of our Phase 2 assessment for the sewage collection models in Table 6.8.

Table 6.8: Phase 2 assessment results for the sewage collection models

Driver	Baseline (PR19 SWC1)	SWC1 Model 2	Baseline (PR19 SWC2)	SWC2 Model 2	SWC2 Model 6	SWC2 Model 8
Sewer length (log)	0.804*** {0.000}	0.842*** {0.000}	0.859*** {0.000}	0.867*** {0.000}	0.847*** {0.000}	0.852*** {0.000}
Pumping capacity per sewer length (log)	0.344** {0.012}	0.357** {0.017}	0.604*** {0.000}	0.568*** {0.000}	0.594*** {0.000}	0.554*** {0.000}
Properties per sewer length (log)	1.043*** {0.000}	0.972*** {0.000}				
LAD weighted average density (log)			-2.480** {0.021}	-2.102*** {0.000}		
Square LAD weighted average density (log)			0.181*** {0.010}	0.158*** {0.000}		
Urban rainfall per sewer length (log)		0.116*** {0.000}		0.155*** {0.000}		
MSOA weighted average density, aggregated by LAD (log)					-2.291** {0.041}	
Square MSOA weighted average density, aggregated by LAD (log)					0.169** {0.021}	
MSOA weighted average density (log)						-5.051* {0.060}
Square MSOA weighted average density (log)						0.336** {0.039}
Constant	-7.956*** {0.000}	-7.760*** {0.000}	3.606 {0.395}	2.51 {0.245}	3.016 {0.501}	14.241 {0.195}
R-squared	0.917	0.920	0.895	0.917	0.897	0.895
RESET test	0.356	0.170	0.269	0.670	0.326	0.399
VIF (max)	2.337	2.535	1.930[a]	1.931[a]	1.914[a]	1.996[a]
Pooling	0.720	0.898	0.988	0.973	0.982	0.987

Driver	Baseline (PR19 SWC1)	SWC1 Model 2	Baseline (PR19 SWC2)	SWC2 Model 2	SWC2 Model 6	SWC2 Model 8
Normality	0.394	0.085	0.268	0.001	0.244	0.376
Heteroskedasticity	0.299	0.282	0.051	0.002	0.034	0.027
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000	0.000	0.000
Minimum efficiency score	0.910	0.934	0.874	0.890	0.877	0.856
Maximum efficiency score	1.127	1.17	1.206	1.185	1.206	1.157
Sensitivity of estimated coefficients to removal of most and least efficient company	Green	Green	Green	Green	Green	Amber
Sensitivity of estimated coefficients to removal of first and last year of sample	Green	Amber	Green	Green	Green	Green

Source: CEPA analysis

[a] The reported VIF excludes the squared density term

***Key**: Green Amber Red*

In summary, our statistical tests show that:

- All estimated coefficients are significant and have the expected sign. The linear MSOA density term is only significant at the 10% level when included in the SWC2 model.
- All models have a high R^2, and a low spread of efficiency scores. The inclusion of the normalised urban rainfall drivers produces the highest explanatory power in both sewage collection models, as well as the smallest efficiency score spread.
- The models all perform well on the RESET, Pooling and VIF tests. The LM test for all models is supportive of a Random Effects specification.
- Of the six models we ran, three fail the normality test, and four fail the heteroskedasticity test. We do not consider the results of these tests to be problematic, due to the use of a Random Effects modelling approach and robust standard errors, respectively.

Overall, the conclusions of our Phase 2 assessment of SWC models are as follows.

The **normalised urban rainfall** variable performs well against statistical tests:

- The variable is highly statistically significant in both SWC models. The inclusion of the normalised urban rainfall drivers produces the highest explanatory power in both sewage collection models, as well as the smallest efficiency score spread.
- However, the point estimates and significance of both urban rainfall drivers are substantially reduced when we remove the first year of data (2011-12). For example, when we remove the first year of data from the SWC1 model with urban rainfall per sewer length, the point estimate falls from 0.116 (p=0.000) to 0.056 (p=0.207). This may raise concerns for the overall stability of the driver. This finding appears to be unique to the urban rainfall drivers and may be driven by year-on-year volatility of the urban rainfall data. Indeed, 2011-12 was the driest year in the sample - industry-wide annual rainfall was 7,401 mm, 25% below the average across the whole period. A potential solution that Ofwat might explore is testing a smoothed urban rainfall variable.
- The introduction of the urban rainfall drivers has a positive efficiency impact on NWT and WSH, and a negative efficiency impact on ANH, NES and TMS. Only NWT is impacted by over 10pps by the driver.

We also consider that the variable has a strong engineering rationale as a proxy for drainage inflows. Overall, we recommend that Ofwat should consider the introduction of the normalised urban rainfall variable in the SWC models for PR24 as this improves the explanatory power of the model, especially for the SWC2 model. However, further investigation of the sensitivity of the estimated coefficient to changes in the sample period should be carried out once an additional year of data becomes available.

We find that there are no improvements in the statistical performance of the SWC2 model using the **alternative density variables**:

- The explanatory power of the models with alternative density variables is similar to that of the PR19 models and the statistical significance of the coefficients for the MSOA density variable is also slightly lower than that of the PR19 density variables.
- The alternative density drivers are not robust to the removal of companies from the sample. When we run the SWC2 model with the alternative density drivers on the full dataset the linear and quadratic terms are statistically significant. They are of the expected sign, with a negative linear and positive quadratic term which is indicative of economies of scale.
    - When we remove WSX (as the most efficient company) from the SWC2 model with the MSOA density driver, both the linear and quadratic density drivers become statistically insignificant, but point estimates remain of the expected sign. The introduction of the MSOA density driver improves the efficiency score of YKY by 5pps and worsens scores for WSH by 4pps.

- We note that the results found above are expected, and not necessarily indicative of a poor model specification, given that WSX is the least densely populated company.

The findings are similar to those for the water models using alternative density variables. Compared to the PR19 measure, the MSOA density drivers have the advantage of not relying on companies' mapping of LADs to company areas dataset which improves the transparency and reliability of the dataset. We would also expect the choice of density variables to be consistent across different models. For example, if the MSOA density measure is chosen then it would be applied instead of the LAD measure in both water and wastewater models. Our overall conclusion on the alternative density variables remains the same as set out in Section 6.1.

6.2.2. Sewage treatment models

We present the full results of our Phase 2 assessment for the sewage treatment models in Table 6.9 and Table 6.10.

Table 6.9: Phase 2 assessment results for SWT1 models

Driver	Baseline (PR19 SWT1)	SWT1 Model 4	SWT1 Model 11	SWT1 Model 12
Load (log)	0.653*** {0.000}	0.777*** {0.000}	0.833*** {0.000}	0.748*** {0.000}
Load treated in size bands 1 to 3 (%)	0.029 {0.211}	0.03 {0.145}	0.032* {0.066}	0.034* {0.064}
Load treated with ammonia consent ≤ 3mg/l	0.006*** {0.000}		0.006*** {0.000}	0.005*** {0.000}
Load treated with BOD consent ≤ 7mg/l		0.032*** {0.000}		
Population in coastal areas (%)			0.009** {0.025}	
Pumping capacity per sewer length (log)				0.333* {0.077}
Constant	-3.734*** {0.004}	-5.206*** {0.001}	-6.198*** {0.000}	-6.198*** {0.000}
R-squared	0.854	0.841	0.887	0.891
RESET test	0.056	0.000	0.000	0.025
VIF (max)	5.337	2.724	7.313	5.396
Pooling	0.999	0.992	0.997	1.000
Normality	0.024	0.028	0.333	0.022
Heteroskedasticity	0.417	0.476	0.012	0.014
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000
Minimum efficiency score	0.816	0.736	0.807	0.826
Maximum efficiency score	1.500	1.382	1.243	1.169
Sensitivity of estimated coefficients to removal of most and least efficient company	Amber	Amber	Red	Green
Sensitivity of estimated coefficients to removal of first and last year of sample	Green	Green	Amber	Amber

Source: CEPA analysis

***Key**:* *Green* *Amber* *Red*

Table 6.10: Phase 2 assessment results for SWT2 models

Driver	Baseline (PR19 SWT2)	SWT2 Model 4	SWT2 Model 11	SWT2 Model 12	SWT2 Model 13	SWT2 Model 14	SWT2 Model 15	SWT2 Model 16	SWT2 Model 17
Load (log)	0.658*** {0.000}	0.783*** {0.000}	0.890*** {0.000}	0.736*** {0.000}	0.723*** {0.000}	0.669*** {0.000}	0.713*** {0.000}	0.686*** {0.000}	0.665*** {0.000}
Load treated with ammonia consent ≤ 3mg/l	0.006*** {0.000}		0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}
Load treated with BOD consent ≤ 7mg/l		0.034*** {0.000}							
Population in coastal areas (%)			0.011*** {0.003}						
Pumping capacity per sewer length (log)				0.322 {0.113}					
Load treated in size band 6 (%)	-0.009* {0.097}	-0.009 {0.102}	-0.012*** {0.002}	-0.010** {0.043}					
Load treated in STWs ≥ 100,000 people (%)					-0.008*** {0.007}				
Load treated in STWs ≥ 125,000 people (%)						-0.005*** {0.008}			
Load treated in STWs ≥ 250,000 people (%)							-0.007*** {0.003}		
Load treated in STWs ≥ 500,000 people (%)								-0.009*** {0.000}	
Load treated in STWs ≥ 1,000,000 people (%)									-0.007*** {0.002}

Driver	Baseline (PR19 SWT2)	SWT2 Model 4	SWT2 Model 11	SWT2 Model 12	SWT2 Model 13	SWT2 Model 14	SWT2 Model 15	SWT2 Model 16	SWT2 Model 17
Constant	-2.965***	-4.444***	-5.931***	-4.051***	-4.072***	-3.567***	-4.142***	-3.848***	-3.730***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
R-squared	0.855	0.841	0.898	0.891	0.869	0.870	0.899	0.919	0.879
RESET test	0.142	0.001	0.045	0.178	0.272	0.221	0.463	0.827	0.389
VIF (max)	4.349	2.096	6.647	4.453	5.347	5.019	4.443	4.663	6.241
Pooling	1.000	0.999	0.998	1.000	1.000	1.000	0.998	0.995	1.000
Normality	0.045	0.099	0.111	0.012	0.221	0.186	0.057	0.234	0.397
Heteroskedasticity	0.875	0.899	0.027	0.083	0.764	0.991	0.927	0.536	0.060
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Minimum efficiency score	0.844	0.753	0.838	0.857	0.875	0.872	0.910	0.921	0.869
Maximum efficiency score	1.505	1.381	1.204	1.197	1.410	1.393	1.357	1.197	1.367
Sensitivity of estimated coefficients to removal of most and least efficient company	−	−	−	−	−	−	+	+	+
Sensitivity of estimated coefficients to removal of first and last year of sample	−	−	+	−	+	+	+	+	+

Source: CEPA analysis

***Key**:* + *Green* − *Amber* × *Red*

Driver	Baseline (PR19 SWT2)	SWT2 Model 18	SWT2 Model 19	SWT2 Model 20
Load (log)	0.658*** {0.000}	0.747*** {0.000}	0.770*** {0.000}	0.788*** {0.000}
Load treated with ammonia consent ≤ 3mg/l	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}
Load treated in size band 6 (%)	-0.009* {0.097}			
Weighted average CEPA size band (log)		-2.015*** {0.000}		
Weighted average ANH size band (log)			-1.485*** {0.000}	
Weighted average treatment size (log)				-0.242*** {0.000}
Constant	-2.965*** {0.000}	-1.091 {0.221}	-2.292*** {0.001}	-3.001*** {0.000}
R-squared	0.855	0.887	0.899	0.911
RESET test	0.142	0.228	0.431	0.849
VIF (max)	4.349	4.545	4.454	4.339
Pooling	1.000	0.999	0.999	0.997
Normality	0.045	0.025	0.021	0.064
Heteroskedasticity	0.875	0.593	0.651	0.865
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000
Minimum efficiency score	0.844	0.899	0.908	0.913
Maximum efficiency score	1.505	1.388	1.354	1.244
Sensitivity of estimated coefficients to removal of most and least efficient company	−	+	+	+
Sensitivity of estimated coefficients to removal of first and last year of sample	−	+	+	+

Source: CEPA analysis

Key: + *Green* − *Amber* × *Red*

In summary, our statistical tests show that:

- Most estimated coefficients are significant and have the expected sign, with the following exceptions:
    - The *load treated in size band 1-3* driver used in the PR19 SWT1 model specification is not significant at the 10% level (p=0.211). This variable becomes significant at the 10% level when the coastal population variable is added.
    - The *load treated in size band 6* in the PR19 SWT1 model specification is also only marginally significant at the 10% level (p=0.097). This may reflect the fact that STW size band 6 covers a wide range of treatment work sizes, and thus a large proportion of companies' sewage treatment.
- Of all sewage treatment models, explanatory power is largest when alternative economies of scale drivers are used in place of the PR19 economies of scale drivers. Use of the coastal driver also significantly improves model explanatory power.
- The models with BOD and coastal population fail the RESET test. All models perform well on the Pooling and VIF tests. The LM test for all models is supportive of a Random Effects specification.
- All models fail either the normality or the heteroskedasticity test. We do not consider this to be problematic, due to the use of a Random Effects modelling approach and the use of robust standard errors.

Overall, the conclusions of our Phase 2 assessment of SWT models are as follows.

We do not consider that the **BOD variable** performs sufficiently well against our statistical criteria to recommend its use at PR24. The introduction of the BOD variable reduces the explanatory power of both SWT1 and SWT2. All models with BOD fail the RESET test. In addition, we noted in Section 5.3 that the materiality of BOD treatment on costs is unclear. Furthermore, 84.5% of all load treated with BOD ≤ 7mg/l was treated by four companies (TMS, SVH, NWT and ANH). It is therefore possible that this driver is picking up company effects that are not related to BOD treatment costs.

The addition of *load treated with BOD consents ≤ 7mg/l* improves efficiency scores for TMS and SRN, and worsens scores for NWT, SVH and YKY.

We do not consider that the **coastal population** variable performs sufficiently well against our statistical criteria to recommend its use at PR24:

- The introduction of the coastal population variable improves the explanatory power of the SWT1 and SWT2 baseline models.
- However, the coastal models are not robust to the exclusion of the company with the largest share of coastal population, as discussed in more detail below.
- All models including the coastal population variable also fail the RESET test.
- We also note that the dataset for coastal populations was provided by Southern Water, and has not yet been verified by Ofwat.

The significance and sign of the coastal driver is sensitive to the removal of SRN (the least efficient company) in both sewage treatment models:

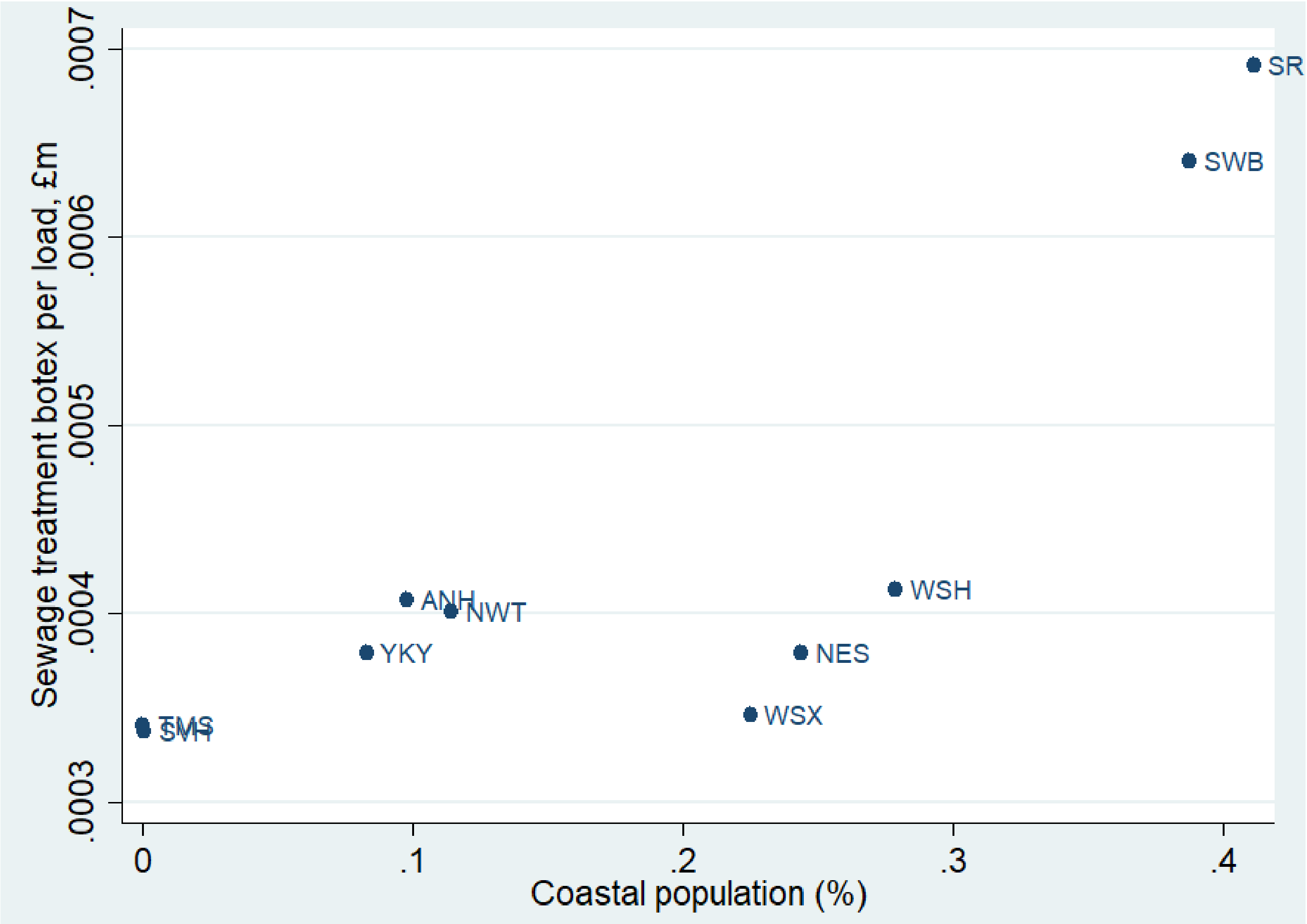
- When we run the SWT1 model with the coastal driver on the full dataset, the driver is significant at the 5% level and has a positive sign, in line with the expectation that sewage treatment costs are higher for companies with more coastal population. However, when we run the model with the coastal driver on the sample excluding SRN, the coefficient of the coastal driver becomes negative, with a point estimate of -0.688 (p=0.645).
- When we run the SWT2 model with the coastal driver on the full dataset, the driver is significant at the 1% level and has a positive sign. However, when we run the model with the coastal driver on the dataset excluding SRN,

the coefficient of the coastal driver becomes statistically insignificant, with a positive point estimate of 0.135 (p=0.921).

We have analysed the relationship between coastal population and costs to understand why the models are sensitive to the removal of one company. Southern Water has over 40% of its residential customers in coastal areas, the highest proportion of the wastewater companies. However, it is not a data outlier, as a similar proportion of South West Water's customer base is also in coastal areas.

As the figure below demonstrates, the likely reason that the significance and sign of the coastal population driver is sensitive to exclusion of companies is because SRN and SWB are data outliers. These companies both have a larger proportion of coastal population, and significantly larger sewage treatment unit costs. This could be due to inefficiency, or alternative cost drivers that are not captured by the coastal variable, or because there is a non-linear relationship between coastal populations and sewage treatment unit costs. The inclusion of the coastal driver improves efficiency scores for SRN by 26 and 30pps in SWT1 and SWT2 respectively, while it worsens the efficiency score of YKY, NES and SVH, which all decrease by 5-10pps.

Figure 6.1: Relationship between coastal populations and sewage treatment costs: 2021-22



Source: CEPA Analysis, Southern Water

We also do not recommend the use of the **pumping capacity per sewer length** driver. The driver was suggested as an alternative driver of costs associated with coastal operation of the sewage treatment system. However, as noted in our Phase 1 assessment, pumping capacity should not include pumps associated with sea outflows, and so the correlation between pumping capacity and the coastal population driver may be spurious. For this reason, there is insufficient rationale for the inclusion of the pumping capacity driver in the SWT base cost models.

We find that the **economies of scale** drivers in the PR19 SWT model specifications have lost statistical significance:

- Load treated in bands 1 to 3, used in the PR19 SWT1 model specification, is no longer statistically significant at the 10% level in SWT1 (p=0.211), and
- Load treated in band 6 is now only significant at the 10% level when used in the PR19 SWT2 model specification (p=0.097). In its PR19 redetermination, the CMA noted that there is significant variation in STW size within STW size band 6.[44]

For these reasons, we believe Ofwat should consider alternative economies of scale drivers for the sewage treatment models at PR24, where supported by economic and technical rationale.

The alternative sewage treatment economies of scale drivers all perform well statistically and offer strong alternatives to the PR19 economies of scale drivers:

- The two **weighted average STW size band** drivers, using our 8-band structure and Anglian Water's 10-band structure, as well as the **Weighted Average Treatment Size** driver, perform well statistically as alternative economies of scale drivers. Each is significant at the 1% level, and their significance and sign are robust to the removal of data. The model explanatory power is stronger than the driver for load treated in STWs ≥ 100,000 people.
- Based on engineering input, we understand that STWs that serve more than 100,000 people are typically large enough to provide additional processes such as sludge treatment. Therefore, there may be changes in costs associated with treatment works of this size. Both **Load treated in STWs ≥ 100,000** and **Load treated in STWs ≥ 125,000 people** perform strongly in the sewage treatment models. Load treated in STWs ≥ 125,000 people is however sensitive to the removal of Thames Water (as a data outlier and the most efficient company under this model); the removal of Thames Water increases the p-value of the driver from 0.008 to 0.252, though the sign remains negative as expected a priori. **Load treated in STWs ≥ 100,000 people** is less sensitive to the removal of companies. When TMS is removed, the variable is still significant at the 10% level (with a p-value of 0.069).
- **Load treated in STWs ≥ 250,000 people** also performs well statistically (SWT2 Model 15). The driver is significant (p=0.003), the model provides a large model explanatory power (R^2 = 0.899), and is robust to the removal of data. Load treated in STWs ≥ 250,000 people accounts for 56.8% of total load in 2021-22, however this is comprised of only 41 STWs (compared to 128 STWs for the 100,000 threshold). The econometric results for this driver could become skewed by a smaller number of observations.
- **Load treated in STWs ≥ 500,000 people** is the strongest statistically performing of the economies of scale drivers (SWT2 Model 16). The driver is statistically significant (p=0.000), provides the largest model explanatory power (R^2 = 0.919), and narrows the range of efficiency scores. The sign and significance of this driver is also robust to the removal of data. One concern with this variable is that the result may be driven by few STWs, leading to the increased likelihood of results being driven by outliers. Load treated in STWs ≥ 500,000 people accounts for 29.9% of the total load in 2021-22, though this is comprised of only 17 STWs. Of all load treated in STWs ≥ 500,000 people, three companies account for the majority of load: Thames Water (40.6%), Severn Trent (15.2%) and United Utilities (13.4%). We do note, however, that statistical results are robust to both the removal of Thames Water as an outlier, and the most and least efficient companies.
- **Load treated in STWs ≥ 1,000,000 people** performs well statistically, but we have concerns with regards to the variation of this variable across companies (STW2 Model 17). The driver is significant (p=0.002) and improves model explanatory power relative to the PR19 SWT2 baseline (R^2=0.879). Furthermore, model 17 passes all statistical tests at the 5% significance level, and the results are robust to the removal of companies and years. However, we note that only 16% of all load treated in 2021-22 are treated in STWs ≥ 1,000,000 people, or only 5 STWs. Furthermore, 72.2% of all load treated in STWs ≥ 1,000,000 people is treated by

[44] CMA (2021), Anglian Water Services Limited, Bristol Water plc, Northumbrian Water Limited and Yorkshire Water Services Limited price determinations

Thames Water. Whilst the results of this model specification are robust to the removal of Thames Water, we do not recommend this model specification due to data concerns.[45]

In summary, there are multiple feasible alternatives for the economies of scale variable which statistically outperform the PR19 drivers. Engineering input has indicated load treated in STWs ≥ 100,000 people as a relevant cut-off point for the economies of scale driver. This variable performs more strongly than STWs ≥ 125,000 people which is sensitive to the removal of Thames Water. However, STWs ≥ 100,000 people does not perform as well statistically as load treated in STWs ≥ 250,000 people, load treated in STWs ≥ 500,000 people, the two weighted average band drivers and the weighted average treatment size driver. These variables produce the highest explanatory power and are most robust to removal of observations from the sample. However, as described above, the main concern regarding load treated in STWs ≥ 250,000 people and load treated in STWs ≥ 500,000 people is that the performance of a small handful of companies may drive the results given the reduced number of observations for STWs above 250,000 and 500,000 people respectively.

We consider the CEPA WAB variable to be preferable to the ANH WAB variable based on statistical performance, in the round. In the WWWN+ models, presented in Section 6.2.3, the CEPA WAB variable performs better than the ANH WAB variable from a statistical perspective.

When choosing between the remaining drivers, after filtering for statistical performance, technical rationale and evidence on the relationship between STW, size and costs would ideally be used to support the choice of driver. Engineering input from CEPA and Ofwat has not provided a clear technical understanding of the relationship between STW size and unit costs. Ofwat should further explore the relationship between STW size and unit costs in order to ascertain which driver is more suitable. For example:

- The cut-off drivers suggest a binary relationship between STW size and unit costs. A cut-off driver would be most appropriate if there is a specific size threshold at which the largest step-change in unit costs takes place.
- The WATS variable aims to estimate a simple linear, continuous relationship between STW size and unit costs; if a linear relationship is backed by technical rationale, the WATS variable would be preferable.
- The WAB drivers derive a more complicated relationship between STW sizes across companies and unit costs. A WAB driver would be most appropriate if potential opportunities for economies of scale were unlocked at different size thresholds.

Transparency, simplicity and robustness of the assumptions underlying the variable calculation are also important factors to consider when selecting a cost driver. We consider the WATS variable to be easy to understand. The interpretation of the WATS variable is simple, tying the weighted average size of STWs to costs. However, a limitation of this variable is the lack of a complete dataset for all STWs which means that the calculation of the WATS variable involves two distinct approaches for STWs in bands 1-5, and STWs in bands above 5. In contrast, the WAB variable is less easy to interpret given the use of assumed bands to separate STWs. Furthermore, the results are more dependent on the thresholds chosen to define the bands.

In conclusion, we suggest that there are three candidate economies of scale drivers that merit further consideration from Ofwat; load treated in STWs ≥ 100,000 people, the CEPA weighted average band, and weighted average treatment size. These drivers perform well statistically, are unlikely to be skewed by the results of a few STWs, and can be justified on technical grounds.

[45] When we run SWT2 Model 17 without Thames Water, the driver for STWs ≥ 1,000,000 people has a point estimate of -0.006, with a p-value of 0.000. The other drivers remain significant and of the expected sign. However, the model's explanatory power falls substantially, to R^2=0.857. Model explanatory power under this specification does not fall as sharply when the most and least efficient company are removed, nor when we remove years of data from the sample.

6.2.3. Wastewater network plus models

We present the full results of our Phase 2 assessment for the wholesale wastewater network plus models in Table 6.11.

Table 6.11: Phase 2 assessment results for wholesale wastewater network plus models

Driver	Baseline (WWWN+ model 21)	WWWN+ Model 23	WWWN+ Model 19	WWWN+ Model 25	WWWN+ Model 26	WWWN+ Model 27	WWWN+ Model 28	WWWN+ Model 29	WWWN+ Model 30
Load (log)	0.646*** {0.000}	0.650*** {0.000}	0.727*** {0.000}	0.691*** {0.000}	0.686*** {0.000}	0.706*** {0.000}	0.696*** {0.000}	0.676*** {0.000}	0.617*** {0.000}
Pumping capacity per sewer length (log)	0.367*** {0.000}	0.355*** {0.000}	0.380*** {0.000}	0.370*** {0.001}	0.359*** {0.000}	0.347*** {0.000}	0.330*** {0.002}	0.259*** {0.003}	0.381*** {0.000}
Load treated with ammonia consent ≤ 3mg/l	0.005*** {0.000}	0.005*** {0.000}	0.005*** {0.000}	0.005*** {0.000}	0.005*** {0.000}	0.005*** {0.000}	0.005*** {0.000}	0.006*** {0.000}	0.005*** {0.000}
Urban rainfall per sewer length (log)		0.076** {0.017}							
Load treated in STWs of size bands 1-3 (%)			0.023* {0.073}						
Load treated in STWs of size band 6 (%)				-0.004 {0.167}					
Load treated in STWs ≥ 100,000 people (%)					-0.002 {0.204}				
Load treated in STWs ≥ 125,000 people (%)						-0.003** {0.021}			
Load treated in STWs ≥ 250,000 people (%)							-0.003 {0.162}		
Load treated in STWs ≥ 500,000 people (%)								-0.004***	

Driver	Baseline (WWWN+ model 21)	WWWN+ Model 23	WWWN+ Model 19	WWWN+ Model 25	WWWN+ Model 26	WWWN+ Model 27	WWWN+ Model 28	WWWN+ Model 29	WWWN+ Model 30
								{0.002}	
Load treated in STWs ≥ 1,000,000 people (%)									0.002 {0.487}
Constant	-2.984*** {0.000}	-2.807*** {0.000}	-4.106*** {0.000}	-3.228*** {0.000}	-3.374*** {0.000}	-3.578*** {0.000}	-3.505*** {0.000}	-3.238*** {0.000}	-2.628*** {0.000}
R-squared	0.947	0.953	0.952	0.949	0.949	0.950	0.954	0.959	0.941
RESET test	0.572	0.163	0.478	0.677	0.700	0.783	0.798	0.918	0.176
VIF (max)	4.169	4.239	5.396	4.453	5.348	5.023	4.471	4.666	6.276
Pooling	0.978	0.997	0.992	0.997	0.997	0.996	0.979	0.922	0.984
Normality	0.435	0.651	0.044	0.101	0.352	0.329	0.246	0.126	0.316
Heteroskedasticity	0.515	0.231	0.603	0.762	0.333	0.325	0.196	0.101	0.089
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Minimum efficiency score	0.917	0.934	0.911	0.927	0.921	0.918	0.928	0.949	0.878
Maximum efficiency score	1.069	1.104	1.082	1.087	1.067	1.070	1.058	1.077	1.097
Sensitivity of estimated coefficients to removal of most and least efficient company	Green	Green	Green	Amber	Amber	Green	Amber	Green	Amber
Sensitivity of estimated coefficients to removal of first and last year of sample	Green	Amber	Green	Green	Amber	Green	Amber	Green	Green

Source: CEPA analysis

***Key**: Green Amber Red*

Driver	Baseline (WWWN+ model 21)	WWWN+ Model 31	WWWN+ Model 32	WWWN+ Model 33
Load (log)	0.646***	0.715***	0.703***	0.714***
	{0.000}	{0.000}	{0.000}	{0.000}
Pumping capacity per sewer length (log)	0.367***	0.332***	0.334***	0.295***
	{0.000}	{0.001}	{0.001}	{0.002}
Load treated with ammonia consent ≤ 3mg/l	0.005***	0.005***	0.005***	0.005***
	{0.000}	{0.000}	{0.000}	{0.000}
Weighted average CEPA size band (log)		-0.843***		
		{0.002}		
Weighted average ANH size band (log)			-0.464	
			{0.158}	
Weighted average treatment size (log)				-0.092**
				{0.012}
Constant	-2.984***	-2.283***	-2.806***	-2.929***
	{0.000}	{0.000}	{0.000}	{0.000}
R-squared	0.947	0.953	0.954	0.956
RESET test	0.572	0.801	0.771	0.901
VIF (max)	4.169	4.545	4.484	4.352
Pooling	0.978	0.991	0.983	0.973
Normality	0.435	0.094	0.092	0.102
Heteroskedasticity	0.515	0.474	0.316	0.167
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000
Minimum efficiency score	0.917	0.939	0.936	0.953
Maximum efficiency score	1.069	1.068	1.059	1.092

Driver	Baseline (WWWN+ model 21)	WWWN+ Model 31	WWWN+ Model 32	WWWN+ Model 33
Sensitivity of estimated coefficients to removal of most and least efficient company	Green	Green	Amber	Green
Sensitivity of estimated coefficients to removal of first and last year of sample	Green	Green	Amber	Green

Source: CEPA analysis

[a] The reported VIF is for the model excluding the squared density term

***Key**: Green Amber Red*

In summary, our statistical tests show that:

- Most estimated coefficients are significant and have the expected sign. But only some of the alternative economies of scale drivers are statistically significant when included in the WWWN+ models.
- All models demonstrate a very high explanatory power. The inclusion of urban rainfall or most of the alternative economies of scale drivers results in a slight improvement in the explanatory power compared to the PR19 model specification.
- All models perform well against the RESET, VIF, pooling, normality and heteroskedasticity tests. The LM test for all models is supportive of a Random Effects specification.

Overall, the conclusions of our Phase 2 assessment of SWT models are as follows.

As for the SWC models, the inclusion of the **normalised urban rainfall** variable improves the explanatory power of the WWWN+ model and should be considered further. As set out before, one potential area of concern is the sensitivity of the estimated coefficient to changes in the sample period which should be investigated further.

For **the economies of scale** drivers in the WWWN+ models, the different STW size band cut-offs considered have a mixed performance when used in the WWWN+ models:

- **Load treated in bands 1 to 3** performs well statistically overall; the variable is significant with a p-value of 0.073, and is of the expected sign. Furthermore, the sign and significance of the driver is robust to the removal of data. However, this driver is not statistically significant at the 10% level in the baseline SWT1 model.
- **Load treated in band 6** does not perform well statistically; the variable is not statistically significant at the 10% significance level, and is not robust to the removal of the least efficient company.
- The best performing drivers statistically in the WWWN+ models are **load treated in STWs ≥ 125,000 and 500,000 people**, which are statistically significant and robust to the removal of data. These models also improve the explanatory power of the models.[46] However, as noted in section 6.2.2, we have concerns that load treated in STWs ≥ 500,000 relies on a smaller number of observations.
- The driver for **load treated in STWs ≥ 100,000 people** is statistically insignificant in the full sample (p=0.204). However, when we remove either the first year of data (2011-12) or WSX (as the most efficient company), the driver becomes statistically significant (p=0.026 for the removal of the first year, and p=0.036 for the removal of WSX).
- The fact that similar economies of scale drivers (such as load treated in STWs ≥ 100,000 and ≥ 125,000 people) have such a mixed performance suggests that the performance of these drivers in the WWWN+ model is sensitive to the threshold selected. We note that, in contrast, the same drivers are consistently performing well in the SWT models; this is likely because STW size predominantly affects sewage treatment costs, and so the relationship between STW size and WWWN+ costs is less pronounced.

Overall, the three weighted treatment size drivers perform reasonably well statistically:

- The **CEPA weighted average band** and the **weighted average treatment size** drivers are statistically significant and of the expected sign. Each improves the explanatory power of the model relative to the baseline, by 6 to 9pps.
- The ANH weighted average band driver is not significant at the 10% level (p-value of 0.158). However, it becomes statistically significant with a p-value of 0.015 when we remove WSX (as the most efficient company), and becomes significant at the 10% level when we remove the first year of data (2011-12).

---

[46] R^2 = 0.947 for the baseline WWWN+ model, increasing to R^2=0.950 and 0.959 when *load treated in STWs ≥ 125,000 and 500,000 people* are included, respectively.

- We note that the point estimate of the pumping capacity per sewer length driver is substantially lower when the weighted average treatment size driver is included; the point estimate falls from 0.367 in the baseline WWWN+ model to 0.295 when the weighted average treatment size driver is included. This suggests some interaction effects between pumping capacity and average STW size. This may be because companies that have smaller STWs are more likely to operate in sparse areas and near the coast, which is likely to require more pumping.

When evaluating the economies of scale drivers, we need to take account of statistical performance across both the sewage treatment and wholesale wastewater network plus models, in addition to the performance of the different variables viz a viz data availability and sample size, transparency, simplicity, and technical rationale. These aspects are discussed in some detail in Section 6.2.2. We discount the load treated in STWs ≥ 500,000 people because the driver is at particular risk of being skewed by the results of a few large STWs.

We discuss in Section 6.2.2 that engineering input has not provided a strong case for any specific economies of scale driver, except for the use of a 100,000 population threshold which is not statistically significant at the 10% level in the WWWN+ models. The STW size band drivers perform well in the STW models but are more sensitive to the chosen threshold in the WWWN+ models. Ofwat should therefore be cautious when considering these variables for inclusion in the WWWN+ models.

Overall, we recommend two economies of scale drivers to Ofwat for consideration in the WWWN+ models: **CEPA's weighted average band,** and the **weighted average treatment size,** as these drivers perform well statistically, and cannot be separated by technical rationale, simplicity or transparency as discussed in Section 6.2.2.

6.3. Bioresources models

Table 6.12 summarises the results of our Phase 2 assessment, which considers the variables we progressed from our Phase 1 assessment, whose findings are summarised in Table 5.3.

Table 6.12: Variables considered in our Phase 2 assessment

Model	BR1	BR2
Alternative density drivers		
MSOA (population)	✖	
LAD from MSOA (population)	⊖	
Properties per length of mains	✖	
Economies of scale in sludge treatment		
CEPA size bands	✖	✖
Weighted average bands (WAB)	✖	⊖

Key: ⊕ *Consider* ⊖ *Needs further consideration* ✖ *Do not consider*

We present the full results of our Phase 2 assessment for the bioresources models in tables Table 6.13 and Table 6.14.

.

Table 6.13: Phase 2 assessment results for the BR1 models

Driver	BR1 Model 1	BR1 Model 2	BR1 Model 3	BR1 Model 4	BR1 Model 11	BR1 Model 14	BR1 Model 15
	PR19	MSOA instead of LAD	LAD from MSOA	Prop. per sewer length instead of LAD	Load treated in STWs >250k	CEPA weighted average band	ANH weighted average band
Sludge produced (log)	1.172***	1.132***	1.176***	1.136***	1.157***	1.121***	1.123***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
WAD - LAD (log)	**-0.133**				**-0.07**	**-0.109**	**-0.096**
	{0.267}				**{0.668}**	**{0.512}**	**{0.576}**
Load treated in STWs bands 1-3 (%)	0.063**	0.064**	0.063**	0.046**			
	{0.011}	{0.016}	{0.011}	{0.045}			
WAD - MSOA (log)		**-0.093**					
		{0.642}					
WAD - LAD from MSOA (log)			**-0.139**				
			{0.217}				
Properties per sewer length (log)				**-0.638**			
				{0.218}			
Load treated in STWs ≥ 250,000 people (%)					-0.010**		
					{0.029}		
Weighted average treatment band (CEPA) (%)						-1.851*	
						{0.064}	
Weighted average treatment band (ANH) (%)							-1.242**
							{0.039}
Constant	-0.912	-0.946	-0.889	0.744	-0.708	2.826*	1.658

Driver	BR1 Model 1	BR1 Model 2	BR1 Model 3	BR1 Model 4	BR1 Model 11	BR1 Model 14	BR1 Model 15
	{0.310}	{0.479}	{0.312}	{0.613}	{0.455}	{0.052}	{0.114}
R-squared	0.82	0.815	0.821	0.812	0.783	0.797	0.791
RESET test	0.528	0.409	0.488	**0.009**	0.326	0.292	0.342
VIF (max)	3.086	3.057	3.066	3.877	2.658	2.659	3.039
Pooling	0.676	0.815	0.753	0.999	0.769	0.811	0.813
Normality	0.267	0.149	0.261	0.124	0.334	0.188	0.242
Heteroskedasticity	0.332	0.212	0.338	0.22	0.521	0.518	0.502
Pooled OLS vs RE (LM test)	0	0	0	0	0	0	0
Minimum efficiency score	0.682	0.67	0.684	0.669	0.718	0.671	0.661
Maximum efficiency score	1.434	1.496	1.436	1.591	1.648	1.609	1.601
Sensitivity of estimated coefficients to removal of most and least efficient company	Amber	Amber	Amber	Amber	Amber	Red	Red
Sensitivity of estimated coefficients to removal of first and last year of sample	Green	Green	Green	Amber	Green	Green	Green

Key: *Green* *Amber* *Red*

Table 6.14: Phase 2 assessment results for BR2 models

Driver	BR2 Model 1	BR2 Model 4	BR2 Model 7	BR2 Model 8
	PR19	**>250k**	**CEPA weighted average band**	**ANH weighted average band**
Sludge produced (log)	1.134***	1.123***	1.080***	1.092***
	{0.000}	{0.000}	{0.000}	{0.000}
STWs per property (log)	**0.275**			
	{0.174}			
Load treated in STWs ≥ 250,000 people (%)		-0.011**		
		{0.024}		
Weighted average treatment band (CEPA) (%)			-2.152**	
			{0.027}	
Weighted average treatment band (ANH) (%)				-1.448**
				{0.019}
Constant	0.808	-1.034*	2.792**	1.507*
	{0.316}	{0.075}	{0.034}	{0.071}
R-squared	0.784	0.781	0.795	0.789
RESET test	0.374	0.32	0.442	0.494
VIF (max)	3.359	2.086	2.151	2.347
Pooling	0.974	0.977	0.99	0.99
Normality	**0.04**	0.195	**0.08**	0.138
Heteroskedasticity	0.757	0.397	0.31	0.373
Pooled OLS vs RE (LM test)	0	0	0	0
Minimum efficiency score	0.597	0.712	0.67	0.661
Maximum efficiency score	1.471	1.713	1.72	1.697

Driver	BR2 Model 1	BR2 Model 4	BR2 Model 7	BR2 Model 8
Sensitivity of estimated coefficients to removal of most and least efficient company	Green	Amber	Green	Green
Sensitivity of estimated coefficients to removal of first and last year of sample	Green	Green	Green	Green

Key: *Green* *Amber* *Red*

In summary, our statistical tests show that:

- None of the density drivers tested in BR1 are statistically significant.
- All alternative Economies of Scale drivers are statistically significant.
- Most alternative model specifications have an explanatory power (R^2) that is lower than the PR19 model specifications. The exceptions are:
  - BR1 Model 3, with LAD from MSOA as an alternative density driver.
  - BR2 Models 4 and 9, with CEPA WAB and ANH WAB as alternative Economies of Scale drivers respectively.
- The models generally perform well on the RESET, VIF, and pooling tests. Only BR2 Model 4 using properties per sewer length as the density driver fails the RESET test. The LM test for all models is supportive of a Random Effects specification.
- Some models fail the normality and heteroskedasticity tests. We do not consider this problematic due to the use of a Random Effects specification and the use of robust standard errors respectively.
- We find that all the BR2 models are robust to the removal of companies and years from the sample.
- We find that some BR1 models are not robust to the removal of companies and years from the sample, we discuss this in more detail below.

We set out our overall conclusions of our Phase 2 assessment of BR models below.

For the **alternative density variables,** we find that none of the density variables are statistically significant. We further tested the BR1 model without the density variable (i.e., we tested a model with only scale and economies of scale drivers). We found that the impact on R-squared was minimal, decreasing from 0.82 for the PR19 specification to 0.817. Therefore, Ofwat could consider dropping the density variable from BR1. However, as the inclusion of a density driver is the main difference between models BR1 and BR2, excluding the density driver could result in Ofwat having to rely on a single model to estimate bioresources costs at PR24. We consider that there is a strong technical rationale as to why density can affect both sludge treatment and sludge transport and disposal costs and this could justify the inclusion of density in the BR models; although the extent to which density captures different impacts from the economies of scale at STWs drivers would need to be explored further.

We also note that the LAD from MSOA variable in model 3 has a lower p-value than the PR19 density variable WAD – LAD in model 1, and the R-squared is marginally higher in model 3 than in model 1. Furthermore, as explained in previous sections, MSOA-based density measures have the advantage, compared to the PR19 density measure, that they do not rely on mapping of company areas to LADs. Therefore, if Ofwat wishes to retain a density driver, we consider that replacing the WAD – LAD variable with the LAD from MSOA variable could represent an improvement particularly in terms of transparency and reliability of the underlying dataset.

For the **alternative economies of scale variables**, we find that the PR19 variable in BR1 (load treated in bands 1-3) performs well. The % of load in STWs ≥ 250,000 people is statistically significant in BR1, but the R-square of this model decreases compared to the PR19 specification. The model is also sensitive to the removal of the most efficient company (NES), in which case the EoS driver becomes statistically insignificant. A further consideration is that the selected economies of scale at STWs driver in the BR models should be consistent with the driver used in the SWT and WWWN+ models. In the case of SWT and WWWN+ models, we concluded that the inclusion of % of load in STWs ≥ 250,000 may not be appropriate as the econometric results could be skewed by the use of a smaller number of observations.

For the BR1 models, the WAB variables are sensitive to the removal of the most efficient company and the last year of data, becoming statistically insignificant. We also note that the density variable in these specifications changes its sign. This is likely to be because we are replacing a variable that accounts for the prevalence of small STWs (generally associated with sparser areas) with a variable that also reflects the impact of large STWs (associated with denser areas). The models including the WAB variables also have a lower R-squared compared to the PR19 model specification. Overall, for the BR1 model, we recommend that Ofwat retains the PR19 variable of Load treated in STWs bands 1-3.

For the BR2 models (that do not include a density driver), we find that the PR19 variable (STWs per property) is statistically insignificant, while the WAB variables and the alternative band size variable are statistically significant. However, the alternative size band (% of load in STWs ≥ 250,000 people) is sensitive to the removal of the most efficient company, in which case the EoS driver becomes statistically insignificant. The WAB variables are robust to the removal of years and companies. Furthermore, we find that the CEPA WAB variable has a higher explanatory power than the ANH WAB variable. One concern related to the WAB variable, as presented in our discussion of SWT models, is that it relies on assumptions about the appropriate definition of bands.

Overall, we suggest that Ofwat consider models that include:

- a scale driver (sludge produced);
- an economies of scale driver: load treated in size bands 1-3, sewage treatment works per number of properties or the CEPA WAB variable; and
- potentially a density driver (with LAD from MSOA being the best performing density variable in the BR models based on our analysis) if there is a sufficiently strong technical rationale to suggest density should be used either in addition or instead of an economies of scale driver.

CEPA economics matters

7. DEVELOPING OPTIONS FOR CONSIDERATION AT PR24

In Section 6 we set out the outcome of our assessment of alternative models that Ofwat might use at PR24. In this section, we summarise the individual variables that we consider to be viable alternatives for the PR24 cost models and further assess their impacts by testing combinations of the alternative variables that performed best based on our analysis.

7.1. Water models

Following the Phase 2 assessment set out in Section 1, we suggest that Ofwat should consider the following variables for the PR24 base cost water models:

- **Average Pumping Head (TWD)** (in the TWD and WW models) – which leads to small improvements in the statistical performance of the TWD model and is supported by a strong technical rationale. APH (TWD) performs slightly better in the WW models in terms of statistical robustness and data quality than APH (all), as discussed in 6.1.3. Therefore, we have used APH (TWD) in the WW models presented in this section as our preferred option.
- **Alternative density drivers** (used at all levels of cost aggregation) – these do not necessarily lead to improvements in the statistical performance of the models but can be justified on the grounds of improving the data quality and the accuracy of the density measure relative to the PR19 density measure which is more prone to error. The recommendations we considered are:
    - MSOA density;
    - LAD density aggregated from MSOA; and
    - properties per length.
- **Length of mains** (in the WW models) – which performs well as an alternative scale driver in the WW model although does not offer a clear improvement over the PR19 model specification and has an impact on the performance of the booster per length driver in the WW models, as discussed in section 6.1.3.

Our assessment shows that all these options perform well against our selection criteria when assessed as an individual change to the PR19 models. However, as many of these decisions affect the same models, Ofwat will also need to consider the combined impact of the options on its base cost models. To help Ofwat understand these potential impacts, we have tested the following combinations of options and assessed the robustness of the model specifications:

- For TWD, we have tested APH (TWD) instead and alongside boosters per length in combination with alternative density drivers.
- For WW, we have tested APH (TWD) instead and alongside boosters per length in combination with alternative density drivers. We also tested an alternative scale driver (length of mains) in combination with APH (TWD) alongside and instead of boosters per length.

We considered the statistical performance and robustness of these additional model specifications by applying the same statistical sensitivity tests set out in Section 1.

We present below a final selection of water models for consideration, including models which test the following different combinations of options in order to assess the robustness of the model specifications.

7.1.1. Water resources plus models

We present the conclusions of our analysis of the WRP model specifications in Section 6.1.1. There are no alternative model specifications or combinations of options that we propose should be included in the base cost modelling suite at PR24. We summarise our final proposed WRP models in Table 7.1 below.

Table 7.1: Final WRP models

Driver	WRP1 Model 3	WRP1 Model 4	WRP1 Model 5	WRP2 Model 3	WRP2 Model 4	WRP2 Model 5
	MSOA instead of LAD density	LAD from MSOA density	Prop. per length instead of LAD density	MSOA instead of LAD density	LAD from MSOA density	Prop. per length instead of LAD density
Properties (log)	1.054***	1.077***	1.028***	1.057***	1.075***	1.027***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.004***	0.005***	0.005***			
	{0.009}	{0.002}	{0.001}			
Weighted average complexity (log)				0.315	0.343	0.365
				{0.234}	{0.183}	{0.143}
Density variable (log)	-4.986**	-1.545***	-7.815**	-5.048**	-1.468**	-7.440**
	{0.017}	{0.007}	{0.019}	{0.034}	{0.026}	{0.030}
Squared density variable (log)	0.303**	0.097***	0.858**	0.306**	0.091**	0.810**
	{0.017}	{0.008}	{0.028}	{0.033}	{0.031}	{0.042}
Constant	9.416	-5.335***	6.988	9.591	-5.660***	6.136
	{0.226}	{0.000}	{0.309}	{0.286}	{0.002}	{0.389}
R-squared	0.901	0.909	0.91	0.896	0.902	0.905
RESET test	0.765	0.436	0.324	0.729	0.367	0.203
VIF (max)	1.269	1.206	1.112	1.308	1.253	1.158
Pooling	1.000	0.999	0.983	1.000	0.999	0.997
Normality	0.417	0.522	0.143	0.416	0.812	0.527
Heteroskedasticity	0.000	0.000	0.000	0.000	0.000	0.000
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000	0.000	0.000
Minimum efficiency score	0.493	0.527	0.507	0.473	0.501	0.484

Driver	WRP1 Model 3	WRP1 Model 4	WRP1 Model 5	WRP2 Model 3	WRP2 Model 4	WRP2 Model 5
Maximum efficiency score	1.997	2.016	1.975	1.983	1.986	1.948
Sensitivity of estimated coefficients to removal of most and least efficient company	+	+	+	−	+	+
Sensitivity of estimated coefficients to removal of first and last year of sample	+	+	+	+	+	+

7.1.2. Treated water distribution and wholesale water models

Density and APH in TWD and WW Models

As both APH (TWD) and alternative density drivers are good alternatives for PR24, we have combined these options for TWD and WW models. The results show good statistical performance when combining MSOA/LAD from MSOA/properties per length with APH alongside or instead of boosters per length in the TWD models (see Table 7.2 below). All variables are statistically significant and pass all robustness tests. The alternative density drivers perform less well in the WW models without boosters per length, when combining APH (TWD) with MSOA/LAD from MSOA, as the water treatment complexity drivers (Water treated in bands 3-6 (%) and WAC) lose some significance or even marginally become statistically insignificant in the model including the MSOA density driver (see Table 7.3 below). In the WW models including boosters per length, all drivers are statistically significant, when combining APH (TWD) and alternative density variables. If Ofwat decides to use the MSOA density driver, we consider that the models including boosters per length are more statistically robust. However, if Ofwat decides to drop boosters per length in the WW models, and the results remain the same with an additional year of data, it may need to reconsider whether the statistical performance of the model is sufficiently robust to include in the PR24 suite of models. Ofwat may still decide to include both APH, boosters per length and MSOA density in the model but in that case, it should reconsider the use of some of the drivers in the model (e.g., explore the inclusion of an alternative driver for treatment complexity).

Table 7.2: Final TWD models

Driver	TWD APH no booster, MSOA	TWD APH + booster, MSOA	TWD APH no booster, LAD from MSOA	TWD APH + booster, LAD from MSOA	TWD APH no booster, prop/length	TWD APH + booster, prop/length
Lengths of main (log)	1.017***	1.017***	1.062***	1.062***	1.045***	1.060***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Density (log)	-6.539***	-5.787***	-2.975***	-2.754***	16.623***	-14.233***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Squared density (log)	0.445***	0.405***	0.229***	0.219***	2.055***	1.811***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Boosters per length (log)		0.305***		0.345***		0.386***
		{0.004}		{0.004}		{0.002}
Average Pumping Head - distribution (log)	0.411***	0.380***	0.357***	0.314***	0.357***	0.339***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Constant	16.573***	14.564***	1.99	2.547**	26.125***	21.933***
	{0.000}	{0.000}	{0.218}	{0.029}	{0.000}	{0.000}
R-squared	0.965	0.968	0.961	0.965	0.966	0.971
RESET test	0.719	0.859	0.439	0.321	0.845	0.856
VIF (max)	1.062	1.599	1.032	1.869	1.037	1.878
Pooling	0.767	0.632	0.798	0.73	0.847	0.536
Normality	0.954	0.469	0.65	0.716	0.474	0.555
Heteroskedasticity	0.828	0.654	0.482	0.673	0.268	0.637
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000	0.000	0.000
Minimum efficiency score	0.708	0.745	0.722	0.751	0.748	0.822
Maximum efficiency score	1.318	1.351	1.315	1.376	1.284	1.377

Driver	TWD APH no booster, MSOA	TWD APH + booster, MSOA	TWD APH no booster, LAD from MSOA	TWD APH + booster, LAD from MSOA	TWD APH no booster, prop/length	TWD APH + booster, prop/length
Sensitivity of estimated coefficients to removal of most and least efficient company	Green	Green	Green	Green	Green	Green
Sensitivity of estimated coefficients to removal of first and last year of sample	Green	Green	Green	Green	Green	Green

Source: CEPA analysis

***Key**:* *Green* *Amber* *Red*

Table 7.3: Final WW models (APH + alternative density variables, no boosters)

Driver	WW1 APH, no booster, MSOA	WW2 APH, no booster, MSOA	WW1 APH, no booster, LAD from MSOA	WW2 APH, no booster, LAD from MSOA	WW1 APH , no booster, prop/length	WW2 APH, no booster, prop/length
Properties (log)	1.041***	1.037***	1.066***	1.059***	1.025***	1.020***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.002*		0.003**		0.003**	
	{0.073}		{0.028}		{0.014}	
Weighted average complexity (log)		0.258		0.290*		0.318**
		{0.108}		{0.075}		{0.036}
Density (log)	-6.145***	-5.895***	-2.179***	-2.036***	-12.767***	-12.007***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Squared density (log)	0.384***	0.367***	0.148***	0.138***	1.467***	1.374***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Average Pumping Head - TWD (log)	0.359***	0.351***	0.345***	0.336***	0.278**	0.265**
	{0.002}	{0.002}	{0.001}	{0.002}	{0.022}	{0.034}
Constant	13.173**	12.138**	-3.750**	-4.293**	16.893***	15.240***
	{0.010}	{0.022}	{0.035}	{0.015}	{0.000}	{0.000}
R-squared	0.961	0.962	0.965	0.965	0.966	0.967
RESET test	0.895	0.935	0.838	0.821	0.781	0.614
VIF (max)	1.271	1.315	1.211	1.267	1.115	1.196
Pooling	0.975	0.97	0.856	0.85	0.979	0.978
Normality	0.395	0.502	0.329	0.794	0.076	0.178
Heteroskedasticity	0.000	0.000	0.000	0.000	0.000	0.000
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000	0.000	0.000

Driver	WW1 APH, no booster, MSOA	WW2 APH, no booster, MSOA	WW1 APH, no booster, LAD from MSOA	WW2 APH, no booster, LAD from MSOA	WW1 APH , no booster, prop/length	WW2 APH, no booster, prop/length
Minimum efficiency score	0.723	0.704	0.745	0.722	0.745	0.726
Maximum efficiency score	1.444	1.431	1.487	1.463	1.453	1.433
Sensitivity of estimated coefficients to removal of most and least efficient company	+	+	+	+	+	−
Sensitivity of estimated coefficients to removal of first and last year of sample	+	+	+	+	−	−

Table 7.4: Final WW models (APH + alternative density variables+ boosters)

Driver	WW1 APH + booster, MSOA	WW2 APH + booster, MSOA	WW1 APH + booster, LAD from MSOA	WW2 APH + booster, LAD from MSOA	WW1 APH + booster, prop/length	WW2 APH + booster, prop/length
Properties (log)	1.047***	1.042***	1.069***	1.061***	1.036***	1.031***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.003**		0.003**		0.003***	
	{0.027}		{0.011}		{0.006}	
Weighted average complexity (log)		0.287**		0.321**		0.334**
		{0.050}		{0.034}		{0.019}
Density (log)	-5.015***	-4.748***	-1.931***	-1.769***	-10.787***	-10.044***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Squared density (log)	0.321***	0.303***	0.136***	0.124***	1.263***	1.171***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Boosters per length (log)	0.420***	0.409***	0.357**	0.354***	0.322**	0.306**
	{0.003}	{0.003}	{0.011}	{0.009}	{0.030}	{0.028}
Average Pumping Head - TWD (log)	0.298***	0.285**	0.285***	0.270**	0.249**	0.234*
	{0.007}	{0.012}	{0.007}	{0.015}	{0.037}	{0.062}
Constant	10.078***	8.962***	-3.207***	-3.814***	13.447***	11.812***
	{0.002}	{0.010}	{0.008}	{0.002}	{0.003}	-10.044***
R-squared	0.969	0.970	0.970	0.971	0.969	0.970
RESET test	0.964	0.580	0.645	0.545	0.703	0.400
VIF (max)	1.79	1.744	1.96	1.952	1.902	1.888
Pooling	0.995	0.99	0.978	0.963	0.974	0.971
Normality	0.846	0.874	0.969	0.974	0.388	0.395

Driver	WW1 APH + booster, MSOA	WW2 APH + booster, MSOA	WW1 APH + booster, LAD from MSOA	WW2 APH + booster, LAD from MSOA	WW1 APH + booster, prop/length	WW2 APH + booster, prop/length
Heteroskedasticity	0.000	0.000	0.000	0.000	0.000	0.000
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000	0.000	0.000
Minimum efficiency score	0.782	0.758	0.774	0.749	0.759	0.740
Maximum efficiency score	1.361	1.379	1.338	1.357	1.321	1.306
Sensitivity of estimated coefficients to removal of most and least efficient company	Amber	Amber	Green	Green	Amber	Green
Sensitivity of estimated coefficients to removal of first and last year of sample	Green	Amber	Green	Green	Green	Amber

Source: CEPA analysis

Key: Green Amber Red

Length of mains in the WW models

Including an additional WW model with length of mains as scale driver, instead of number of properties, could be considered for PR24. We tested this model in combination with LAD from MSOA and with and without APH (TWD) (alongside or instead of boosters per length). All variables are statistically significant on a 10% basis and pass all robustness tests, apart from the heteroskedasticity test, which does not cause any issues due to the use of robust standard errors (see Table 7.5 below). Boosters per length loses some statistical significance, when combining with length per mains and, even more so, when APH (TWD) is also included.

Ofwat may consider using other alternative density drivers (MSOA or properties per length) in the WW models using length of mains. However, we consider that the choice of density driver is a separate decision and this should not affect the choice of whether or not to use length of mains in the WW models. Therefore, to limit the number of models presented here, we have only included results for model specifications using the LAD from MSOA density driver (as this is the density driver that most resembles the PR19 WAD LAD measure).

Table 7.5: Combined WW models (Length of mains + LAD from MSOA)

Driver	WW1 Length of mains, LAD from MSOA, booster	WW2 Length of mains, LAD from MSOA, booster	WW1 Length of mains, LAD from MSOA, APH, booster	WW2 Length of mains, LAD from MSOA, APH, booster	WW1 Length of mains, LAD from MSOA, APH, no booster	WW2 Length of mains, LAD from MSOA, APH, no booster
Lengths of main (log)	1.047***	1.037***	1.045***	1.037***	1.044***	1.037***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.003***		0.003***		0.003**	
	{0.001}		{0.007}		{0.013}	
Weighted average complexity (log)		0.370**		0.330**		0.319**
		{0.011}		{0.033}		{0.048}
WAD - MSOA to LAD (log)	-2.255***	-2.063***	-2.347***	-2.179***	-2.474***	-2.315***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Squared WAD - MSOA to LAD (log)	0.179***	0.164***	0.184***	0.172***	0.190***	0.178***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Boosters per length (log)	0.288**	0.279**	0.210*	0.210*		
	{0.032}	{0.023}	{0.093}	{0.080}		
APH - TWD (log)			0.239**	0.219**	0.271***	0.254***
			{0.011}	{0.028}	{0.004}	{0.008}
Constant	2.631*	1.77	1.677	1.033	1.296	0.671
	{0.053}	{0.198}	{0.204}	{0.441}	{0.286}	{0.600}
R-squared	0.968	0.970	0.970	0.972	0.968	0.969
RESET test	0.155	0.094	0.485	0.407	0.663	0.642
VIF (max)	1.835	1.834	1.884	1.882	1.095	1.194
Pooling	0.871	0.728	0.918	0.85	0.918	0.897
Normality	0.632	0.434	0.441	0.259	0.069	0.144
Heteroskedasticity	0	0	0.001	0	0.005	0.004

Driver	WW1 Length of mains, LAD from MSOA, booster	WW2 Length of mains, LAD from MSOA, booster	WW1 Length of mains, LAD from MSOA, APH, booster	WW2 Length of mains, LAD from MSOA, APH, booster	WW1 Length of mains, LAD from MSOA, APH, no booster	WW2 Length of mains, LAD from MSOA, APH, no booster
Pooled OLS vs RE (LM test)	0	0	0	0	0	0
Minimum efficiency score	0.774	0.772	0.819	0.795	0.803	0.779
Maximum efficiency score	1.355	1.354	1.428	1.4	1.518	1.49
Sensitivity of estimated coefficients to removal of most and least efficient company	Amber	Green	Amber	Green	Green	Green
Sensitivity of estimated coefficients to removal of first and last year of sample	Green	Green	Green	Green	Green	Green

Source: CEPA analysis

***Key**:* *Green* *Amber* *Red*

7.2. WASTEWATER MODELS

Following the Phase 2 assessment set out in Section 1, we suggest that Ofwat should consider the following variables for the PR24 base cost wastewater models:

- **Normalised urban rainfall** (in the SWC and WWWN+ models) – as this results in overall improvements in the performance of the models.
- **Alternative density drivers** (in the SWC2 model) - these do not lead to improvements in the statistical performance of the models but can be justified on the grounds of improving the data quality and the accuracy of the density measure. The options we considered are:
  - MSOA density; and
  - LAD density aggregated from MSOA.[47]
- **Alternative economies of scale drivers for sewage treatment** – as this can improve the performance of the models given that the PR19 drivers have low or no statistical significance.
  - For the SWT models, we put forward three alternative economies of scale drivers; load treated in STWs ≥ 100,000people, CEPA's weighted average band, and weighted average treatment size.
  - For the WWWN+ models, we consider 2 economies of scale drivers; CEPA's weighted average band and weighted average treatment size.

Based on our assessment in Section 1, we found that all of these options perform well against our selection criteria when assessed as an individual change to the PR19 models. In this section, we present a final selection of models for consideration, including models which test the following different combinations of options in order to assess the robustness of the model specifications. These model results are presented in the tables below.

- For SWC1, we only considered one additional driver to the PR19 baseline specification, namely the normalised urban rainfall driver. We do not test any combination of model specifications in this section.
- For SWC2, we have tested urban rainfall in combination with MSOA density and LAD from MSOA.
- For the SWT models, we have only recommended a selection of the economies of scale drivers. Therefore, we do not test any combination of model specifications in this section.
- For WWWN+, we have tested urban rainfall in combination with CEPA's weighted average band, and the weighted average treatment size driver.

We have considered the statistical performance and robustness of these additional model specifications by applying the same statistical sensitivity tests set out in Section 1.

We present the related efficiency scores associated with each of our final model recommendations in Appendix B.

[47] At PR19, Ofwat used properties per length of sewers as a density driver in its SWC1 model specification. We do not re-test this driver for the SWC2 model here.

Table 7.6: Final SWC1 models

Driver	Baseline (PR19 SWC1)	SWC1 Model 2
Sewer length (log)	0.804*** {0.000}	0.842*** {0.000}
Pumping capacity per sewer length (log)	0.344** {0.012}	0.357** {0.017}
Properties per sewer length (log)	1.043*** {0.000}	0.972*** {0.000}
Urban rainfall per sewer length (log)		0.116*** {0.000}
Constant	-7.956*** {0.000}	-7.760*** {0.000}
R-squared	0.917	0.920
RESET test	0.356	0.170
VIF (max)	2.337	2.535
Pooling	0.720	0.898
Normality	0.394	0.085
Heteroskedasticity	0.299	0.282
Pooled OLS vs RE (LM test)	0.000	0.000
Minimum efficiency score	0.910	0.934
Maximum efficiency score	1.127	1.17
Sensitivity of estimated coefficients to removal of most and least efficient company	+	+
Sensitivity of estimated coefficients to removal of first and last year of sample	+	−

Table 7.7: Final SWC2 models

Driver	SWC2 Model 6 MSOA	SWC2 Model 8 LAD from MSOA	SWC2 Model 13 Urban rainfall + MSOA	SWC2 Model 14 Urban rainfall + LAD from MSOA
Sewer length (log)	0.847*** {0.000}	0.852*** {0.000}	0.865*** {0.000}	0.857*** {0.000}
Pumping capacity per sewer length (log)	0.594*** {0.000}	0.554*** {0.000}	0.517*** {0.000}	0.554*** {0.000}
MSOA Weighted average density (log)	-2.291** {0.041}		-4.693*** {0.005}	
Square MSOA Weighted average density (log)	0.169** {0.021}		0.315*** {0.002}	
MSOA Weighted average density, aggregated by LAD (log)		-5.051* {0.060}		-1.989*** {0.001}
Square MSOA Weighted average density, aggregated by LAD (log)		0.336** {0.039}		0.151*** {0.000}
Urban rainfall per sewer length (log)			0.156*** {0.000}	0.157*** {0.000}
Constant	3.016 {0.501}	14.241 {0.195}	13.056** {0.048}	2.196 {0.325}
R-squared	0.897	0.895	0.918	0.920
RESET test	0.326	0.399	0.759	0.630
VIF (max)	1.914[a]	1.996[a]	2.00[a]	1.92[a]
Pooling	0.982	0.987	0.967	0.951
Normality	0.244	0.376	0.003	0.001
Heteroskedasticity	0.000	0.000	0.002	0.003
Minimum efficiency score	0.877	0.856	0.877	0.897
Maximum efficiency score	1.206	1.157	1.13	1.182

Driver	SWC2 Model 6 MSOA	SWC2 Model 8 LAD from MSOA	SWC2 Model 13 Urban rainfall + MSOA	SWC2 Model 14 Urban rainfall + LAD from MSOA
Sensitivity of estimated coefficients to removal of most and least efficient company	Green	Amber	Green	Green
Sensitivity of estimated coefficients to removal of first and last year of sample	Green	Green	Green	Green

Source: CEPA analysis

[a] The reported VIF excludes the squared density term

***Key**: Green Amber Red*

Table 7.8: Final SWT models

Driver	SWT2 Model 13	SWT2 Model 18	SWT2 Model 20
Load (log)	0.723*** {0.000}	0.747*** {0.000}	0.788*** {0.000}
Load treated with ammonia consent ≤ 3mg/l	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}
Load treated in STWs ≥ 100,000 people (%)	-0.008*** {0.007}		
Weighted average CEPA size band (log)		-2.015*** {0.000}	
Weighted average treatment size (log)			-0.242*** {0.000}
Constant	-4.072*** {0.000}	-1.091 {0.221}	-3.001*** {0.000}
R-squared	0.869	0.887	0.911
RESET test	0.272	0.228	0.849
VIF (max)	5.347	4.545	4.339
Pooling	1.000	0.999	0.997
Normality	0.221	0.025	0.064
Heteroskedasticity	0.764	0.593	0.865
Pooled OLS vs RE (LM test)	0.000	0.000	0.000
Minimum efficiency score	0.875	0.899	0.913
Maximum efficiency score	1.410	1.388	1.244
Sensitivity of estimated coefficients to removal of most and least efficient company	−	+	+
Sensitivity of estimated coefficients to removal of first and last year of sample	+	+	+

Table 7.9: Final WWWN+ models

Driver	Baseline (WWWN+ model 21)	WWWN+ Model 23 Urban rainfall	WWWN+ Model 31 CEPA WAB	WWWN+ Model 33 WATS	WWWN+ Model 44 Urban rainfall + CEPA WAB	WWWN+ Model 45 Urban rainfall + CEPA WATS
Load (log)	0.646***	0.650***	0.715***	0.714***	0.725***	0.722***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Pumping capacity per sewer length (log)	0.367***	0.355***	0.332***	0.295***	0.316***	0.271***
	{0.000}	{0.000}	{0.001}	{0.002}	{0.000}	{0.000}
Load treated with ammonia consent ≤ 3mg/l	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Urban rainfall per sewer length (log)		0.076**			0.089**	0.092***
		{0.017}			{0.011}	{0.009}
Weighted average CEPA size band (log)			-0.843***		-0.885***	
			{0.002}		{0.003}	
Weighted average treatment size (log)				-0.092**		-0.096***
				{0.012}		{0.001}
Constant	-2.984***	-2.807***	-2.283***	-2.929***	-2.068***	-2.712***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
R-squared	0.947	0.953	0.953	0.956	0.962	0.964
RESET test	0.572	0.163	0.801	0.901	0.097	0.318
VIF (max)	4.169	4.239	4.545	4.352	4.550	4.469
Pooling	0.978	0.997	0.991	0.973	0.932	0.909
Normality	0.435	0.651	0.094	0.102	0.249	0.242
Heteroskedasticity	0.515	0.231	0.474	0.167	0.093	0.067
Pooled OLS vs RE (LM test)	0.000	0.000	0.000	0.000	0.000	0.001

Driver	Baseline (WWWN+ model 21)	WWWN+ Model 23 Urban rainfall	WWWN+ Model 31 CEPA WAB	WWWN+ Model 33 WATS	WWWN+ Model 44 Urban rainfall + CEPA WAB	WWWN+ Model 45 Urban rainfall + CEPA WATS
Minimum efficiency score	0.917	0.934	0.939	0.953	0.956	0.970
Maximum efficiency score	1.069	1.104	1.068	1.092	1.060	1.055
Sensitivity of estimated coefficients to removal of most and least efficient company	+	+	+	+	+	+
Sensitivity of estimated coefficients to removal of first and last year of sample	+	−	+	+	−	−

7.2.1. Sewage collection models

Both combined SWC2 models perform well, in line with our Phase 2 assessment of the individual options. All estimated coefficients are statistically significant at the 1% level and there is a notable improvement in the explanatory power of the model when the urban rainfall driver is included. In particular, it is worth noting that the performance of the alternative density drivers (especially MSOA) improves when urban rainfall is also included in the SWC2 model.

7.2.2. Sewage treatment models

We present the conclusions of our analysis of the alternative economies of scales drivers in Section 6.2.2. There are no alternative drivers that we propose should be included in the amended base cost modelling suite.

7.2.3. Wastewater network plus models

The models including urban rainfall and weighted average economies of scale drivers perform well. All estimated coefficients are statistically significant at the 1% level and there is a notable improvement in the explanatory power of the model compared to the baseline WWWN+ model. The point estimate of the urban rainfall and the two economies of scale drivers in the combination model specifications are comparable to the model specifications considered in Section 6.2.3.[48]

However, the significance and point estimate of the urban rainfall driver remains sensitive to the removal of the first year of data (2011-12). In the model with urban rainfall and the CEPA weighted average band, the point estimate of the urban rainfall variable falls from 0.089 under the full sample (p=0.011) to 0.059 (p=0.220) when the first year of data is excluded. Similarly, in the model with urban rainfall and weighted average treatment size, the point estimate of the urban rainfall variable falls from 0.092 under the full sample (p=0.009) to 0.063 (p=0.182) when the first year of data is excluded.

Given the strength of these drivers individually and in combination, these WWWN+ models could be considered for PR24.

7.3. Bioresources models

Following the Phase 2 assessment set out in Section 1, we suggest that Ofwat should consider the following variables for the PR24 base cost water models:

- **Density drivers:** we would recommend either excluding the density driver from the BR models, as none of the density drivers tested are statistically significant, or using the LAD from MSOA variable as this is the driver that performs best from a statistical perspective in the BR1 model.
- **Economies of scale at SWT drivers:** the following variables can be considered:
    - % load treated in bands 1-3 (the PR19 BR1 EoS driver);
    - STWs per property (the PR19 BR2 EoS driver); and
    - CEPA weighted average band.

We present these models below. The only addition compared to the models presented in Phase 2 is a model excluding density drivers. The best performing model in terms of explanatory power remains the PR19 BR1 model specification using LAD from MSOA as an alternative model specification. Removing the density driver slightly

[48] The point estimate for the urban rainfall driver is similar in WWWN+ Model 23, compared to the model specifications including urban rainfall and either economies of scale driver, as shown in Table 7.7. The point estimate for the CEPA weighted average band is -0.843 in WWWN+ Model 31 (p=0.002), and is -0.885 when the normalised urban rainfall driver is also added (p=0.003). Similarly, the point estimate for the weighted average treatment size is -0.092 in WWWN+ Model 33 (p=0.012), and is -0.096 when the normalised urban rainfall driver is also added (p=0.001).

reduces the explanatory power of the model but also slightly improves the significance of the economies of scale driver.

Table 7.10: Final bioresources models

Driver	BR1 Model 3	BR1 Model 26	BR2 Model 1	BR1 Model 7
	LAD from MSOA	No density variable	PR19 BR2	CEPA weighted average band
Sludge produced (log)	1.176*** {0.000}	1.119*** {0.000}	1.134*** {0.000}	1.080*** {0.000}
WAD - LAD from MSOA (log)	**-0.139** **{0.217}**			
Load treated in STWs bands 1-3 (%)	0.063** {0.011}	0.073*** {0.004}		
STWs per property (log)			**0.275** **{0.174}**	
Weighted average treatment band (CEPA) (%)				-2.152** {0.027}
Constant	-0.889 {0.312}	-1.654** {0.014}	0.808 {0.316}	2.792** {0.034}
R-squared	0.821	0.817	0.784	0.795
RESET test	0.488	0.278	0.374	0.442
VIF (max)	3.066	2.455	3.359	2.151
Pooling	0.753	0.944	0.974	0.99
Normality	0.261	0.141	**0.04**	**0.08**
Heteroskedasticity	0.338	0.197	0.757	0.31
Pooled OLS vs RE (LM test)	0	0	0	0
Minimum efficiency score	0.684	0.675	0.597	0.67
Maximum efficiency score	1.436	1.527	1.471	1.72
Sensitivity of estimated coefficients to removal of most and least efficient company	−	−	+	+
Sensitivity of estimated coefficients to removal of first and last year of sample	+	+	+	+

7.4. Summary of recommendations

For the PR24 base cost water modelling, we recommend that Ofwat should consider the following models:

- The PR19 treated water distribution models and wholesale water models with the inclusion of **APH** (TWD) if the APH data is sufficiently robust once an additional year of data becomes available. Ofwat should also consider whether there is an engineering rationale for including boosters per length alongside APH.

- An alternative wholesale water model with **length of mains** used as the scale driver, though the impact on the boosters per length variable of including length of mains and possibly APH in the WW models should be considered further.

For all water models, Ofwat should consider whether an **alternative density driver** (MSOA, LAD aggregated from MSOA or properties per length of mains) would improve data quality and increase the robustness of the density variable. If an alternative density driver is chosen, then Ofwat should review how the alternative density driver performs in combination with any other changes to the model specification.

For the PR24 base cost wastewater modelling, we recommend that Ofwat should consider the following models:

- The PR19 sewage collection models with the addition of **normalised urban rainfall.** Ofwat should also consider whether an **alternative density driver** (MSOA or LAD aggregated from MSOA) would improve the data quality and robustness of the density variable.
- The PR19 sewage treatment models with **alternative economies of scale** drivers. Ofwat should consider which alternative driver is best supported by technical rationale. Our assessment indicates that the drivers that perform the best against our model selection criteria are:
    - For economies of scale drivers based on threshold size (SWT models only):
        - load treated in STWs ≥ 100,000.
    - For weighted average economies of scale drivers (SWT and WWWN+ models):
        - CEPA's weighted average band.
        - Weighted Average Treatment Size
- Wastewater network plus models that include:
    - load;
    - pumping capacity per sewer length;
    - load treated with ammonia consent ≤ 3mg/l;
    - urban rainfall per sewer length; and
    - a weighted average band or weighted average treatment size variable.

For the PR24 **base cost bioresources models**, we recommend that Ofwat should consider models that include:

- a scale driver (sludge produced);
- an economies of scale driver: load treated in size bands 1-3, sewage treatment works per number of properties or the CEPA WAB variable; and
- potentially a density driver (with LAD from MSOA being the best performing density variable in the BR models based on our analysis) if there is a sufficiently strong technical rationale to suggest density should be used either in addition or instead of an economies of scale driver.

CEPA economics matters

Appendix A PHASE 1 RESULTS

A.1. WATER MODELS

A.1.1. Average Pumping Head

Table A-1: WRP1 with Average Pumping Head

Model	WRP1 Model 1	WRP1 Model 2
Version	PR19	APH (all but distribution), no booster stations
Properties (log)	1.074***	1.099***
	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.006***	0.005**
	{0.000}	{0.010}
WAD - LAD (log)	-1.614***	-1.889***
	{0.000}	{0.000}
Squared WAD - LAD (log)	0.101***	0.122***
	{0.000}	{0.000}
APH - transport, resources and treatment (log)		0.152
		{0.209}
Constant	-5.093***	-5.076***
	{0.000}	{0.000}
R-squared	0.917	0.919
RESET test	0.439	0.745

Table A-2: WRP2 with Average Pumping Head

Model	WRP2 Model 1	WRP2 Model 2
Version	PR19	APH (all but distribution), no booster stations
Properties (log)	1.069***	1.098***
	{0.000}	{0.000}
Weighted average complexity (log)	0.377	0.281
	{0.123}	{0.241}
WAD - LAD (log)	-1.412***	-1.753***
	{0.005}	{0.001}
Squared WAD - LAD (log)	0.087***	0.112***
	{0.009}	{0.001}
APH - transport, resources and treatment (log)		0.165
		{0.112}
Constant	-5.805***	-5.618***
	{0.000}	{0.000}
R-squared	0.907	0.911
RESET test	0.324	0.584

Table A-3: TWD with Average Pumping Head

Model	TWD Model 1	TWD Model 2	TWD Model 3
Version	**PR19**	**APH (distribution), no booster stations**	**APH (distribution), with booster stations**
Lengths of main (log)	1.077***	1.069***	1.069***
	{0.000}	{0.000}	{0.000}
Boosters per length (log)	0.437***		0.333***
	{0.002}		{0.008}
WAD - LAD (log)	-2.946***	-3.203***	-2.879***
	{0.000}	{0.000}	{0.000}
Squared WAD - LAD (log)	0.235***	0.245***	0.228***
	{0.000}	{0.000}	{0.000}
APH - distribution (log)		0.313***	0.276***
		{0.000}	{0.000}
Constant	4.723***	2.892*	3.024**
	{0.002}	{0.057}	{0.032}
R-squared	0.957	0.96	0.964
RESET test	0.102	0.599	0.476

Table A-4: WW1 with Average Pumping Head

Model	WW1 Model 1	WW1 Model 2	WW1 Model 3	WW1 Model 4
Version	**PR19**	**APH (all), no booster stations**	**APH (distribution), no booster stations**	**APH (all), with booster stations**
Properties (log)	1.071***	1.096***	1.066***	1.093***
	{0.000}	{0.000}	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.004***	0.003*	0.004***	0.003**
	{0.000}	{0.059}	{0.005}	{0.032}
WAD - LAD (log)	-2.094***	-2.579***	-2.321***	-2.250***
	{0.000}	{0.000}	{0.000}	{0.000}
Squared WAD - LAD (log)	0.147***	0.177***	0.157***	0.159***
	{0.000}	{0.000}	{0.000}	{0.000}
Boosters per length (log)	0.335**			0.318**
	{0.032}			{0.017}
APH - distribution (log)			0.279***	
			{0.004}	
APH - all (log)		0.323***		0.297***
		{0.008}		{0.007}
Constant	-1.565*	-2.888*	-3.028***	-2.838***
	{0.074}	{0.072}	{0.010}	{0.005}
R-squared	0.97	0.969	0.971	0.972
RESET test	0.223	0.786	0.827	0.597

Table A-5: WW2 with Average Pumping Head

Model	WW2 Model 1	WW2 Model 2	WW2 Model 3	WW2 Model 4
Version	PR19	APH (all), no booster stations	APH (dist), no booster stations	APH (all), with booster stations
Properties (log)	1.059***	1.087***	1.057***	1.083***
	{0.000}	{0.000}	{0.000}	{0.000}
Boosters per length (log)	0.334**			0.319**
	{0.019}			{0.012}
WAD - LAD (log)	-1.832***	-2.401***	-2.104***	-2.051***
	{0.000}	{0.000}	{0.000}	{0.000}
Squared WAD - LAD (log)	0.128***	0.164***	0.142***	0.145***
	{0.000}	{0.000}	{0.000}	{0.000}
Weighted average complexity (log)	0.430***	0.280*	0.371**	0.309*
	{0.001}	{0.094}	{0.021}	{0.052}
APH - distribution (log)			0.270***	
			{0.009}	
APH - all (log)		0.309***		0.275**
		{0.008}		{0.012}
Constant	-2.589***	-3.486**	-3.833***	-3.476***
	{0.001}	{0.021}	{0.001}	{0.000}
R-squared	0.971	0.969	0.97	0.973
RESET test	0.122	0.824	0.771	0.543

A.1.2. Density

Table A-6: WRP1 with alternative density measures

Model	WRP1 Model 1	WRP1 Model 3	WRP1 Model 4	WRP1 Model 5	WRP1 Model 6
Version	PR19	MSOA instead of LAD	LAD from MSOA	Prop. per length instead of LAD	No squared density term
Properties (log)	1.074***	1.054***	1.077***	1.028***	1.045***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.006***	0.004***	0.005***	0.005***	0.006***
	{0.000}	{0.009}	{0.002}	{0.001}	{0.000}
WAD - LAD (log)	-1.614***				-0.199**
	{0.000}				{0.025}
Squared WAD - LAD (log)	0.101***				
	{0.000}				
WAD - MSOA (log)		-4.986**			
		{0.017}			
Squared WAD - MSOA (log)		0.303**			
		{0.017}			
WAD - MSOA to LAD (log)			-1.545***		
			{0.007}		
Squared WAD - MSOA to LAD (log)			0.097***		
			{0.008}		
Properties per length (log)				-7.815**	
				{0.019}	
Squared properties per length (log)				0.858**	
				{0.028}	
Constant	-5.093***	9.416	-5.335***	6.988	-9.548***
	{0.000}	{0.226}	{0.000}	{0.309}	{0.000}
R-squared	0.917	0.901	0.909	0.91	0.912
RESET test	0.439	0.765	0.436	0.324	0.6

Table A-7: WRP2 with alternative density measures

Model	WRP2 Model 1	WRP2 Model 3	WRP2 Model 4	WRP2 Model 5	WRP2 Model 6
Version	PR19	MSOA instead of LAD	LAD from MSOA	Prop. per length instead of LAD	No squared density term
Properties (log)	1.069***	1.057***	1.075***	1.027***	1.043***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
WAD - LAD (log)	-1.412***				-0.201**
	{0.005}				{0.030}
Squared WAD - LAD (log)	0.087***				
	{0.009}				
Weighted average complexity (log)	0.377	0.315	0.343	0.365	0.397
	{0.123}	{0.234}	{0.183}	{0.143}	{0.104}
WAD - MSOA (log)		-5.048**			
		{0.034}			
Squared WAD - MSOA (log)		0.306**			
		{0.033}			
WAD - MSOA to LAD (log)			-1.468**		
			{0.026}		
Squared WAD - MSOA to LAD (log)			0.091**		
			{0.031}		
Properties per length (log)				-7.440**	
				{0.030}	
Squared properties per length (log)				0.810**	
				{0.042}	
Constant	-5.805***	9.591	-5.660***	6.136	-9.639***
	{0.000}	{0.286}	{0.002}	{0.389}	{0.000}
R-squared	0.907	0.896	0.902	0.905	0.905
RESET test	0.324	0.729	0.367	0.203	0.302

Table A-8: TWD with alternative density measures

Model	TWD Model 1	TWD Model 4	TWD Model 5	TWD Model 6	TWD Model 7
Version	PR19	MSOA instead of LAD	LAD from MSOA	Prop. per length instead of LAD	No squared density term
WAD - LAD (log)	-2.946***				0.394***
	{0.000}				{0.000}
Squared WAD - LAD (log)	0.235***				
	{0.000}				
Lengths of main (log)	1.077***	1.026***	1.070***	1.072***	1.006***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Boosters per length (log)	0.437***	0.433***	0.461***	0.488***	0.658***
	{0.002}	{0.001}	{0.002}	{0.001}	{0.000}
WAD - MSOA (log)		-5.561***			
		{0.000}			
Squared WAD - MSOA (log)		0.393***			
		{0.000}			
WAD - MSOA to LAD (log)			-2.729***		
			{0.000}		
Squared WAD - MSOA to LAD (log)			0.219***		
			{0.000}		
Properties per length (log)				-14.921***	
				{0.000}	
Squared properties per length (log)				1.898***	
				{0.000}	
Constant	4.723***	15.638***	4.155***	25.065***	-5.385***
	{0.002}	{0.002}	{0.008}	{0.000}	{0.000}
R-squared	0.957	0.952	0.955	0.958	0.926
RESET test	0.102	0.122	0.09	0.489	0.019

Table A-9: WW1 with alternative density measures

Model	WW1 Model 1	WW1 Model 5	WW1 Model 6	WW1 Model 7	WW1 Model 8
Version	PR19	MSOA instead of LAD	LAD from MSOA	Prop. per length instead of LAD	No squared density term
Properties (log)	1.071***	1.052***	1.072***	1.044***	1.035***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.004***	0.003**	0.003***	0.003***	0.003***
	{0.000}	{0.011}	{0.002}	{0.001}	{0.002}
WAD - LAD (log)	-2.094***				0.006
	{0.000}				{0.933}
Squared WAD - LAD (log)	0.147***				
	{0.000}				
Boosters per length (log)	0.335**	0.509***	0.457***	0.377**	0.496***
	{0.032}	{0.003}	{0.008}	{0.033}	{0.004}
WAD - MSOA (log)		-4.684***			
		{0.001}			
Squared WAD - MSOA (log)		0.301***			
		{0.000}			
WAD - MSOA to LAD (log)			-1.849***		
			{0.000}		
Squared WAD - MSOA to LAD (log)			0.132***		
			{0.000}		
Properties per length (log)				-11.259***	
				{0.000}	
Squared properties per length (log)				1.318***	
				{0.000}	
Constant	-1.565*	10.300*	-1.958	15.655***	-7.711***
	{0.074}	{0.056}	{0.206}	{0.003}	{0.000}
R-squared	0.97	0.963	0.965	0.965	0.955
RESET test	0.223	0.178	0.164	0.205	0.286

Table A-10: WW2 with alternative density measures

Model	WW2 Model 1	WW2 Model 5	WW2 Model 6	WW2 Model 7	WW2 Model 8
Version	PR19	MSOA instead of LAD	LAD from MSOA	Prop. per length instead of LAD	No squared density term
Properties (log)	1.059***	1.046***	1.061***	1.036***	1.027***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
WAD - LAD (log)	-1.832***				-0.007
	{0.000}				{0.910}
Squared WAD - LAD (log)	0.128***				
	{0.000}				
Weighted average complexity (log)	0.430***	0.322**	0.354**	0.366***	0.367**
	{0.001}	{0.030}	{0.016}	{0.007}	{0.023}
Boosters per length (log)	0.334**	0.486***	0.444***	0.351**	0.466***
	{0.019}	{0.003}	{0.005}	{0.033}	{0.002}
WAD - MSOA (log)		-4.308***			
		{0.002}			
Squared WAD - MSOA (log)		0.276***			
		{0.001}			
WAD - MSOA to LAD (log)			-1.648***		
			{0.001}		
Squared WAD - MSOA to LAD (log)			0.117***		
			{0.000}		
Properties per length (log)				-10.322***	
				{0.000}	
Squared properties per length (log)				1.201***	
				{0.000}	
Constant	-2.589***	8.674	-2.795*	13.516***	-7.931***
	{0.001}	{0.108}	{0.064}	{0.008}	{0.000}
R-squared	0.971	0.965	0.967	0.968	0.959
RESET test	0.122	0.075	0.075	0.072	0.294

A.1.3. Alternative water treatment complexity drivers

Table A-11: WRP1 With alternative water treatment complexity drivers

Model	WRP1 Model 1	WRP1 Model 7	WRP1 Model 8	WRP1 Model 9	WRP1 Model 10	WRP1 Model 11	WRP1 Model 12	WRP1 Model 13
Version	PR19	Treated water complexity bands 4-6	Treated water complexity bands 5-6	Treated water complexity bands 6	GW treated bands 3-6, SW treated bands 3-6	GW treated bands 3-6, SW treated bands 4-6	GW treated bands 3-6, SW treated bands 5-6	GW treated bands 3-6, SW treated bands 6
Properties (log)	1.074*** {0.000}	1.099*** {0.000}	1.074*** {0.000}	1.098*** {0.000}	1.097*** {0.000}	1.089*** {0.000}	1.093*** {0.000}	1.098*** {0.000}
Water treated in bands 3-6 (%)	0.006*** {0.000}							
WAD - LAD (log)	-1.614*** {0.000}	-1.359** {0.015}	-1.545** {0.021}	-1.740*** {0.008}	-1.601*** {0.001}	-1.652*** {0.001}	-1.604*** {0.003}	-1.613*** {0.001}
Squared WAD - LAD (log)	0.101*** {0.000}	0.081** {0.031}	0.096** {0.030}	0.110** {0.010}	0.100*** {0.003}	0.104*** {0.002}	0.100*** {0.006}	0.101*** {0.003}
Water treated in bands 4-6 (%)		0.004 {0.156}						
Water treated in bands 5-6 (%)			0.002 {0.519}					
Water treated in band 6 (%)				-0.115*** {0.000}				
GW treated in bands 3-6 (%)					0.003** {0.025}	0.003** {0.020}	0.003** {0.028}	0.003** {0.025}

SW treated in bands 3-6 (%)					0.006*** {0.003}			
SW treated in bands 4-6 (%)						-0.001 {0.837}		
SW treated in bands 5-6 (%)							{0.952}	
SW treated in band 6 (%)								-0.067*** {0.000}
Constant	-5.093*** {0.000}	-5.990*** {0.001}	-4.888*** {0.006}	-4.505** {0.011}	-5.744*** {0.000}	-4.815*** {0.006}	-5.064*** {0.001}	-5.079*** {0.000}
R-squared	0.917	0.902	0.893	0.892	0.905	0.909	0.907	0.908
RESET test	0.439	0.492	0.358	0.625	0.55	0.575	0.53	0.62

Table A-12: WW1 with alternative water treatment complexity drivers

Model	WW1 Model 1	WW1 Model 9	WW1 Model 10	WW1 Model 11	WW1 Model 12	WW1 Model 13	WW1 Model 14	WW1 Model 15
Version	PR19	Treated water complexity bands 4-6	Treated water complexity bands 5-6	Treated water complexity bands 6	GW treated bands 3-6, SW treated bands 3-6	GW treated bands 3-6, SW treated bands 4-6	GW treated bands 3-6, SW treated bands 5-6	GW treated bands 3-6, SW treated bands 6
Properties (log)	1.071***	1.089***	1.058***	1.083***	1.092***	1.095***	1.091***	1.086***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.004***							
	{0.000}							
WAD - LAD (log)	-2.094***	-1.796***	-1.871***	-2.108***	-2.201***	-2.083***	-2.139***	-2.192***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Squared WAD - LAD (log)	0.147***	0.124***	0.130***	0.149***	0.152***	0.144***	0.147***	0.152***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Boosters per length (log)	0.335**	0.355**	0.332**	0.402**	0.223*	0.251*	0.271*	0.247*
	{0.032}	{0.030}	{0.036}	{0.038}	{0.075}	{0.054}	{0.066}	{0.056}
Water treated in bands 4-6 (%)		0.003***						
		{0.008}						
Water treated in bands 5-6 (%)			0.003*					
			{0.061}					
Water treated in band 6 (%)				0.057				
				{0.274}				
GW treated in bands 3-6 (%)					0.004***	0.004***	0.003***	0.004***
					{0.000}	{0.000}	{0.000}	{0.000}

SW treated in bands 3-6 (%)					0.006*** {0.000}			
SW treated in bands 4-6 (%)						0.001 {0.362}		
SW treated in bands 5-6 (%)							0.001 {0.442}	
SW treated in band 6 (%)								0.033* {0.058}
Constant	-1.565* {0.074}	-2.573** {0.021}	-1.849** {0.042}	-1.119 {0.396}	-2.300*** {0.002}	-2.149** {0.012}	-1.734** {0.018}	-1.606** {0.031}
R-squared	0.97	0.966	0.965	0.958	0.972	0.972	0.972	0.972
RESET test	0.223	0.133	0.045	0.113	0.346	0.281	0.153	0.279

Table A-13: WRP2 with alternative water treatment complexity drivers

Model	**WRP2 Model 1**	**WRP2 Model 7**	**WRP2 Model 8**	**WRP2 Model 9**
Version	**PR19**	**Alternative WAC 1 (1, 2, 2, 3, 4, 4, 5)**	**Alternative WAC 2 (1, 4, 9, 16, 25, 36, 49)**	**Alternative WAC 3 (1, 4, 4, 9, 16, 16, 25)**
Properties (log)	1.069***	1.075***	1.079***	1.072***
	{0.000}	{0.000}	{0.000}	{0.000}
Weighted average complexity (log)	0.377			
	{0.123}			
WAD - LAD (log)	-1.412***	-1.402***	-1.420***	-1.420***
	{0.005}	{0.007}	{0.008}	{0.008}
Squared WAD - LAD (log)	0.087***	0.086**	0.087**	0.087**
	{0.009}	{0.011}	{0.014}	{0.014}
Alternative WAC 1 (log)		0.430**		
		{0.011}		
Alternative WAC 2 (log)			0.258**	
			{0.035}	
Alternative WAC 3 (log)				0.238
				{0.225}
Constant	-5.805***	-5.882***	-5.990***	-5.967***
	{0.000}	{0.000}	{0.000}	{0.001}
R-squared	0.907	0.906	0.904	0.904
RESET test	0.324	0.381	0.389	0.313

Table A-14: WW2 with alternative water treatment complexity drivers

Model	WW2 Model 1	WW2 Model 9	WW2 Model 10	WW2 Model 11
Version	PR19	Alternative WAC 1 (1, 2, 2, 3, 4, 4, 5)	Alternative WAC 2 (1, 4, 9, 16, 25, 36, 49)	Alternative WAC 3 (1, 4, 4, 9, 16, 16, 25)
Properties (log)	1.059***	1.068***	1.071***	1.059***
	{0.000}	{0.000}	{0.000}	{0.000}
Weighted average complexity (log)	0.430***			
	{0.001}			
WAD - LAD (log)	-1.832***	-1.838***	-1.843***	-1.812***
	{0.000}	{0.000}	{0.000}	{0.000}
Squared WAD - LAD (log)	0.128***	0.129***	0.129***	0.126***
	{0.000}	{0.000}	{0.000}	{0.000}
Boosters per length (log)	0.334**	0.357**	0.357**	0.329**
	{0.019}	{0.017}	{0.021}	{0.022}
Alternative WAC 1 (log)		0.422***		
		{0.001}		
Alternative WAC 2 (log)			0.251***	
			{0.008}	
Alternative WAC 3 (log)				0.283***
				{0.002}
Constant	-2.589***	-2.497**	-2.635**	-2.904***
	{0.001}	{0.010}	{0.011}	{0.001}
R-squared	0.971	0.968	0.967	0.97
RESET test	0.122	0.134	0.128	0.1

A.1.4. Alternative scale drivers in wholesale water models

Table A-15: WW1 with alternative scale drivers

Model	WW1 Model 1	WW1 Model 16
Version	PR19	Lengths of mains instead of nr of properties
Properties (log)	1.071***	
	{0.000}	
Water treated in bands 3-6 (%)	0.004***	0.004***
	{0.000}	{0.000}
WAD - LAD (log)	-2.094***	-2.446***
	{0.000}	{0.000}
Squared WAD - LAD (log)	0.147***	0.192***
	{0.000}	{0.000}
Lengths of main (log)		1.052***
		{0.000}
Boosters per length (log)	0.335**	0.244*
	{0.032}	{0.081}
Constant	-1.565*	3.058**
	{0.074}	{0.031}
R-squared	0.97	0.969
RESET test	0.223	0.152

Table A-16: WW2 with alternative scale drivers

Model	WW2 Model 1	WW2 Model 12
Version	PR19	Lengths of mains instead of nr of properties
Properties (log)	1.059***	
	{0.000}	
WAD - LAD (log)	-1.832***	-2.222***
	{0.000}	{0.000}
Squared WAD - LAD (log)	0.128***	0.175***
	{0.000}	{0.000}
Weighted average complexity (log)	0.430***	0.383***
	{0.001}	{0.006}
Lengths of main (log)		1.042***
		{0.000}
Boosters per length (log)	0.334**	0.244*
	{0.019}	{0.053}
Constant	-2.589***	2.127
	{0.001}	{0.113}
R-squared	0.971	0.97
RESET test	0.122	0.084

CEPA economics matters

A.1.5. Network reinforcement drivers

Table A-17: TWD with network reinforcement drivers

Model	TWD Model 1	TWD Model 8	TWD Model 9	TWD Model 10
Version	PR19	% new properties as proportion of total properties	% properties growth	% population growth
Lengths of main (log)	1.077***	1.074***	1.074***	1.076***
	{0.000}	{0.000}	{0.000}	{0.000}
Boosters per length (log)	0.437***	0.449***	0.449***	0.442***
	{0.002}	{0.001}	{0.001}	{0.002}
WAD - LAD (log)	-2.946***	-2.931***	-2.931***	-2.977***
	{0.000}	{0.000}	{0.000}	{0.000}
Squared WAD - LAD (log)	0.235***	0.233***	0.233***	0.237***
	{0.000}	{0.000}	{0.000}	{0.000}
New properties as a proportion of total properties (%)		0.039		
		{0.337}		
Growth in properties (%)			0.039	
			{0.338}	
Population growth (%)				-0.005
				{0.648}
Constant	4.723***	4.739***	4.739***	4.858***
	{0.002}	{0.002}	{0.002}	{0.002}
R-squared	0.957	0.956	0.956	0.956
RESET test	0.102	0.109	0.109	0.104

CEPA economics matters

A.1.6. Weather related drivers

Table A-18: WRP1 with weather related drivers

Model	WRP1 Model 1	WRP1 Model 14	WRP1 Model 15	WRP1 Model 16	WRP1 Model 17	WRP1 Model 18
Version	PR19	Rainfall	Rainfall per length	PET as proxy for temperature	Peak demand	Peak demand per annual DI
Properties (log)	1.074***	1.074***	1.122***	1.075***	1.230***	1.074***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
WAD - LAD (log)	-1.614***	-1.599***	-1.597***	-1.634***	-1.726***	-1.613***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Squared WAD - LAD (log)	0.101***	0.100***	0.099***	0.103***	0.110***	0.101***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Rainfall (log)		0.05				
		{0.466}				
Rainfall per length (log)			0.047			
			{0.471}			
PET (log)				-0.037		
				{0.901}		
Peak DI (log)					-0.154	
					{0.228}	
Peak DI per annual DI (log)						0.068
						{0.783}
Constant	-5.093***	-5.528***	-5.672***	-4.798**	-5.883***	-5.113***
	{0.000}	{0.000}	{0.000}	{0.033}	{0.000}	{0.000}
R-squared	0.917	0.919	0.918	0.917	0.917	0.917
RESET test	0.439	0.457	0.461	0.439	0.475	0.456

Table A-19: WRP2 with weather related drivers

Model	WRP2 Model 1	WRP2 Model 10	WRP2 Model 11	WRP2 Model 12	WRP2 Model 13	WRP2 Model 14
Version	PR19	Rainfall	Rainfall per length	PET as proxy for temperature	Peak demand	Peak demand per annual DI
Properties (log)	1.069***	1.068***	1.118***	1.070***	1.226***	1.072***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Weighted average complexity (log)	0.377	0.409*	0.397*	0.376	0.375	0.369
	{0.123}	{0.078}	{0.093}	{0.110}	{0.126}	{0.137}
WAD - LAD (log)	-1.412***	-1.347***	-1.360***	-1.433***	-1.529***	-1.438***
	{0.005}	{0.005}	{0.005}	{0.005}	{0.006}	{0.005}
Squared WAD - LAD (log)	0.087***	0.082***	0.082***	0.088***	0.095***	0.088***
	{0.009}	{0.008}	{0.009}	{0.009}	{0.009}	{0.009}
Rainfall (log)		0.055				
		{0.407}				
Rainfall per length (log)			0.049			
			{0.443}			
PET (log)				-0.051		
				{0.852}		
Peak DI (log)					-0.155	
					{0.233}	
Peak DI per annual DI (log)						0.067
						{0.790}
Constant	-5.805***	-6.449***	-6.522***	-5.420**	-6.587***	-5.736***
	{0.000}	{0.000}	{0.000}	{0.039}	{0.000}	{0.000}
R-squared	0.907	0.91	0.909	0.907	0.907	0.906
RESET test	0.324	0.355	0.363	0.327	0.382	0.35

Table A-20: TWD with weather related drivers

Model	TWD Model 1	TWD Model 11	TWD Model 12	TWD Model 13	TWD Model 14	TWD Model 15
Version	PR19	Rainfall	Rainfall per length	PET as proxy for temperature	Peak demand	Peak demand per annual DI
WAD - LAD (log)	-2.946***	-2.874***	-2.874***	-2.909***	-2.462***	-2.968***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Squared WAD - LAD (log)	0.235***	0.229***	0.229***	0.231***	0.190***	0.236***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Lengths of main (log)	1.077***	1.074***	1.038***	1.076***	0.623***	1.081***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.002}	{0.000}
Boosters per length (log)	0.437***	0.429***	0.429***	0.425***	0.472***	0.440***
	{0.002}	{0.001}	{0.001}	{0.001}	{0.001}	{0.002}
Rainfall (log)		-0.036				
		{0.532}				
Rainfall per length (log)			-0.036			
			{0.532}			
PET (log)				0.278**		
				{0.011}		
Peak DI (log)					0.458**	
					{0.016}	
Peak DI per annual DI (log)						0.281
						{0.252}
Constant	4.723***	4.764***	4.764***	2.821*	5.097***	4.750***
	{0.002}	{0.010}	{0.010}	{0.061}	{0.001}	{0.003}
R-squared	0.957	0.958	0.958	0.958	0.961	0.957
RESET test	0.102	0.064	0.064	0.174	0.071	0.117

Table A-21: WW1 with weather related drivers

Model	WW1 Model 1	WW1 Model 17	WW1 Model 18	WW1 Model 19	WW1 Model 20	WW1 Model 21
Version	PR19	Rainfall	Rainfall per length	PET as proxy for temperature	Peak demand	Peak demand per annual DI
Properties (log)	1.071***	1.071***	1.068***	1.068***	0.905***	1.071***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.004***	0.004***	0.004***	0.005***	0.004***	0.005***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
WAD - LAD (log)	-2.094***	-2.098***	-2.093***	-2.081***	-1.984***	-2.119***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Squared WAD - LAD (log)	0.147***	0.147***	0.147***	0.145***	0.138***	0.147***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Boosters per length (log)	0.335**	0.336**	0.333**	0.315**	0.334**	0.308**
	{0.032}	{0.033}	{0.032}	{0.031}	{0.032}	{0.025}
Rainfall (log)		0.006				
		{0.905}				
Rainfall per length (log)			-0.002			
			{0.966}			
PET (log)				0.131		
				{0.389}		
Peak DI (log)					0.164	
					{0.258}	
Peak DI per annual DI (log)						0.331*
						{0.097}
Constant	-1.565*	-1.596	-1.556	-2.503***	-0.694	-1.677**
	{0.074}	{0.169}	{0.229}	{0.006}	{0.511}	{0.046}
R-squared	0.97	0.97	0.97	0.971	0.97	0.972
RESET test	0.223	0.232	0.223	0.259	0.165	0.269

Table A-22: WW2 with weather related drivers

Model	WW2 Model 1	WW2 Model 13	WW2 Model 14	WW2 Model 15	WW2 Model 16	WW2 Model 17
Version	PR19	Rainfall	Rainfall per length	PET as proxy for temperature	Peak demand	Peak demand per annual DI
Properties (log)	1.059***	1.060***	1.065***	1.058***	0.895***	1.058***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
WAD - LAD (log)	-1.832***	-1.842***	-1.835***	-1.820***	-1.726***	-1.819***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Squared WAD - LAD (log)	0.128***	0.129***	0.128***	0.127***	0.120***	0.126***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Weighted average complexity (log)	0.430***	0.421***	0.425***	0.445***	0.424***	0.508***
	{0.001}	{0.001}	{0.001}	{0.000}	{0.001}	{0.000}
Boosters per length (log)	0.334**	0.334**	0.335**	0.327**	0.334**	0.320**
	{0.019}	{0.020}	{0.019}	{0.019}	{0.018}	{0.014}
Rainfall (log)		0.017				
		{0.729}				
Rainfall per length (log)			0.006			
			{0.897}			
PET (log)				0.072		
				{0.621}		
Peak DI (log)					0.163	
					{0.257}	
Peak DI per annual DI (log)						0.274
						{0.161}
Constant	-2.589***	-2.677***	-2.641**	-3.115***	-1.710*	-2.809***
	{0.001}	{0.009}	{0.022}	{0.002}	{0.058}	{0.000}
R-squared	0.971	0.97	0.97	0.971	0.971	0.972
RESET test	0.122	0.145	0.13	0.128	0.082	0.13

CEPA
economics matters

A.1.7. Economies of scale in water resources

Table A-23: WRP1 with economies of scale drivers

Model	WRP1 Model 1	WRP1 Model 19	WRP1 Model 20
Version	PR19	Number of sources	Average DI per source
Properties (log)	1.074***	1.046***	1.079***
	{0.000}	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.006***	0.006***	0.006***
	{0.000}	{0.003}	{0.003}
WAD - LAD (log)	-1.614***	-1.593***	-1.612***
	{0.000}	{0.000}	{0.000}
Squared WAD - LAD (log)	0.101***	0.100***	0.102***
	{0.000}	{0.000}	{0.000}
Number of sources (log)		0.03	
		{0.782}	
Average DI per source (log)			-0.054
			{0.606}
Constant	-5.093***	-4.958***	-5.155***
	{0.000}	{0.000}	{0.000}
R-squared	0.917	0.918	0.918
RESET test	0.439	0.373	0.376

Table A-24: WRP2 with economies of scale drivers

Model	WRP2 Model 1	WRP2 Model 15	WRP2 Model 16
Version	PR19	Number of sources	Average DI per source
Properties (log)	1.069***	1.079***	1.072***
	{0.000}	{0.000}	{0.000}
WAD - LAD (log)	-1.412***	-1.443***	-1.433***
	{0.005}	{0.009}	{0.006}
Squared WAD - LAD (log)	0.087***	0.089**	0.088***
	{0.009}	{0.011}	{0.009}
Weighted average complexity (log)	0.377	0.366	0.375
	{0.123}	{0.163}	{0.156}
Number of sources (log)		-0.009	
		{0.945}	
Average DI per source (log)			-0.016
			{0.899}
Constant	-5.805***	-5.764***	-5.753***
	{0.000}	{0.000}	{0.000}
R-squared	0.907	0.906	0.906
RESET test	0.324	0.297	0.282

A.1.8. Economies of scale at water treatment works

Table A-25: WRP1 with economies of scale at water treatment works

Model	WRP1 Model 1	WRP1 Model 21	WRP1 Model 22	WRP1 Model 23	WRP1 Model 24	WRP1 Model 25	WRP1 Model 26
Version	PR19	% DI in band 1	% DI in bands 1-2	% DI in bands 1-3	% DI in bands 8	% DI in bands 7-8	% DI in bands 6-8
Properties (log)	1.074***	1.081***	1.080***	1.081***	1.060***	1.021***	1.056***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.006***	0.004**	0.004**	0.004***	0.005***	0.004***	0.005***
	{0.000}	{0.021}	{0.036}	{0.009}	{0.000}	{0.001}	{0.010}
WAD - LAD (log)	-1.614***	-1.723***	-1.739***	-1.653***	-1.512***	-1.556***	-1.553***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.009}	{0.000}	{0.001}
Squared WAD - LAD (log)	0.101***	0.107***	0.105***	0.101***	0.094**	0.092***	0.094***
	{0.000}	{0.000}	{0.000}	{0.001}	{0.021}	{0.003}	{0.006}
DI in band 1 (%)		-0.03					
		{0.423}					
DI in bands 1-2 (%)			-0.021				
			{0.114}				
DI in bands 1-3 (%)				-0.008**			
				{0.041}			
DI in band 8 (%)					0.001		
					{0.697}		
DI in bands 7-8 (%)						0.005*	
						{0.075}	
DI in bands 6-8 (%)							0.003
							{0.252}
Constant	-5.093***	-4.498***	-4.181***	-4.595***	-5.223***	-4.369***	-4.986***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.001}
R-squared	0.917	0.918	0.92	0.922	0.917	0.917	0.923
RESET test	0.439	0.411	0.099	0.357	0.401	0.22	0.137

Table A-26: WRP2 with economies of scale at water treatment works

Model	WRP2 Model 1	WRP2 Model 17	WRP2 Model 18	WRP2 Model 19	WRP2 Model 20	WRP2 Model 21	WRP2 Model 22
Version	PR19	% DI in band 1	% DI in bands 1-2	% DI in bands 1-3	% DI in bands 8	% DI in bands 7-8	% DI in bands 6-8
Properties (log)	1.069***	1.079***	1.079***	1.078***	1.052***	1.019***	1.052***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
WAD - LAD (log)	-1.412***	-1.595***	-1.636***	-1.523***	-1.299**	-1.443***	-1.394***
	{0.005}	{0.000}	{0.000}	{0.001}	{0.042}	{0.006}	{0.007}
Squared WAD - LAD (log)	0.087***	0.097***	0.098***	0.091***	0.078*	0.083**	0.082**
	{0.009}	{0.001}	{0.001}	{0.003}	{0.074}	{0.022}	{0.021}
Weighted average complexity (log)	0.377	0.241	0.203	0.251	0.347	0.229	0.281
	{0.123}	{0.148}	{0.271}	{0.218}	{0.175}	{0.259}	{0.277}
DI in band 1 (%)		-0.032					
		{0.339}					
DI in bands 1-2 (%)			-0.022*				
			{0.072}				
DI in bands 1-3 (%)				-0.009***			
				{0.009}			
DI in band 8 (%)					0.001		
					{0.607}		
DI in bands 7-8 (%)						0.005*	
						{0.068}	
DI in bands 6-8 (%)							0.004
							{0.229}
Constant	-5.805***	-4.904***	-4.501***	-5.042***	-5.907***	-4.696***	-5.510***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.001}	{0.001}
R-squared	0.907	0.911	0.914	0.918	0.907	0.908	0.917
RESET test	0.324	0.36	0.078	0.326	0.288	0.236	0.111

A.1.9. Water resource drivers

Table A-27: WRP1 with water resource drivers

Model	WRP1 Model 1	WRP1 Model 27	WRP1 Model 28	WRP1 Model 29	WRP1 Model 30	WRP1 Model 31	WRP1 Model 32	WRP1 Model 33
Version	PR19	% of DI from pumped & impounding reservoirs	% of DI from pumped reservoirs	% of DI from impounding reservoirs	Number of pumped reservoirs	Number of impounding reservoirs	Number of reservoirs	% DI from all pumped sources
Properties (log)	1.074***	1.037***	1.076***	1.047***	1.025***	1.058***	0.923***	1.047***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.006***	0.004***	0.005***	0.006***	0.007	0.007**	0.006	0.006***
	{0.000}	{0.004}	{0.005}	{0.000}	{0.184}	{0.021}	{0.122}	{0.000}
WAD - LAD (log)	-1.614***	-1.259***	-0.925*	-1.808***	-1.614***	-2.014***	-1.376**	-1.814***
	{0.000}	{0.003}	{0.073}	{0.000}	{0.008}	{0.001}	{0.014}	{0.000}
Squared WAD - LAD (log)	0.101***	0.077***	0.05	0.117***	0.102**	0.132***	0.093***	0.118***
	{0.000}	{0.005}	{0.160}	{0.000}	{0.015}	{0.001}	{0.008}	{0.000}
DI from reservoirs (%)		0.005***						
		{0.001}						
DI from pumped reservoirs (%)			0.006***					
			{0.006}					
DI from impounding reservoirs (%)				0.004				
				{0.155}				
Number of pumped reservoirs (log)					0.039			
					{0.675}			
Number of impounding reservoirs (log)						0.058**		
						{0.046}		

Number of reservoirs (log)							0.124*	
							{0.077}	
DI from all pumped sources (%)								-0.004
								{0.139}
Constant	-5.093***	-5.936***	-7.415***	-4.195***	-4.563***	-3.808**	-4.641***	-3.809***
	{0.000}	{0.000}	{0.000}	{0.001}	{0.004}	{0.014}	{0.001}	{0.009}
R-squared	0.917	0.909	0.906	0.918	0.901	0.917	0.912	0.918
RESET test	0.439	0.519	0.733	0.29	0.429	0.298	0.28	0.289

Table A-28: WRP2 with water resource drivers

Model	WRP2 Model 1	WRP2 Model 23	WRP2 Model 24	WRP2 Model 25	WRP2 Model 26	WRP2 Model 27	WRP2 Model 28	WRP2 Model 29
Version	PR19	% of DI from pumped & impounding reservoirs	% of DI from pumped reservoirs	% of DI from impounding reservoirs	Number of pumped reservoirs	Number of impounding reservoirs	Number of reservoirs	% DI from all pumped sources
Properties (log)	1.069***	1.032***	1.074***	1.040***	1.006***	1.027***	0.881***	1.039***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
WAD - LAD (log)	-1.412***	-1.122**	-0.743	-1.632***	-1.473**	-1.742***	-1.053**	-1.639***
	{0.005}	{0.021}	{0.154}	{0.001}	{0.026}	{0.006}	{0.020}	{0.001}
Squared WAD - LAD (log)	0.087***	0.068**	0.037	0.105***	0.093**	0.114***	0.072**	0.105***
	{0.009}	{0.033}	{0.296}	{0.001}	{0.039}	{0.009}	{0.014}	{0.001}
Weighted average complexity (log)	0.377	0.258	0.264	0.368	0.068	0.52	0.4	0.368
	{0.123}	{0.324}	{0.297}	{0.138}	{0.902}	{0.373}	{0.520}	{0.138}
DI from reservoirs (%)		0.006***						
		{0.001}						
DI from pumped reservoirs (%)			0.006***					
			{0.005}					
DI from impounding reservoirs (%)				0.004				
				{0.135}				
Nr. of pumped reservoirs (log)					0.06			
					{0.550}			
Nr. of impounding reservoirs (log)						0.075**		
						{0.012}		
Nr. of reservoirs (log)							0.151**	
							{0.017}	

DI from all pumped sources (%)								-0.004
								{0.120}
Constant	-5.805***	-6.392***	-8.033***	-4.797***	-4.374*	-4.662**	-5.418***	-4.368***
	{0.000}	{0.000}	{0.000}	{0.003}	{0.071}	{0.047}	{0.004}	{0.010}
R-squared	0.907	0.9	0.896	0.91	0.884	0.916	0.908	0.91
RESET test	0.324	0.594	0.545	0.2	0.316	0.253	0.293	0.2

Table A-29: WW1 with water resource drivers

Model	WW1 Model 1	WW1 Model 22	WW1 Model 23	WW1 Model 24	WW1 Model 25	WW1 Model 26	WW1 Model 27	WW1 Model 28
Version	PR19	% of DI from pumped & impounding reservoirs	% of DI from pumped reservoirs	% of DI from impounding reservoirs	Number of pumped reservoirs	Number of impounding reservoirs	Number of reservoirs	% DI from all pumped sources
Properties (log)	1.071***	1.051***	1.072***	1.064***	1.069***	1.128***	1.067***	1.064***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.004***	0.003***	0.003**	0.004***	0.005**	0.006***	0.006***	0.004***
	{0.000}	{0.007}	{0.017}	{0.000}	{0.017}	{0.000}	{0.001}	{0.000}
WAD - LAD (log)	-2.094***	-1.826***	-1.515***	-2.158***	-2.212***	-2.031***	-2.250***	-2.160***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Squared WAD - LAD (log)	0.147***	0.129***	0.106***	0.152***	0.156***	0.140***	0.158***	0.152***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Boosters per length (log)	0.335**	0.390**	0.448**	0.327*	0.440**	0.475***	0.324*	0.327*
	{0.032}	{0.038}	{0.021}	{0.055}	{0.022}	{0.000}	{0.051}	{0.055}
DI from reservoirs (%)		0.003***						
		{0.002}						
DI from pumped reservoirs (%)			0.004***					
			{0.002}					
DI from impounding reservoirs (%)				0.001				
				{0.564}				
Nr. of pumped reservoirs (log)					0.017			
					{0.788}			
Nr. of impounding reservoirs (log)						-0.028***		
						{0.004}		
Nr. of reservoirs (log)							0.008	

							{0.820}	
DI from all pumped sources (%)								-0.001
								{0.547}
Constant	-1.565*	-2.085**	-3.140***	-1.309	-0.802	-1.963**	-1.233	-1.205
	{0.074}	{0.014}	{0.003}	{0.191}	{0.503}	{0.023}	{0.221}	{0.266}
R-squared	0.97	0.968	0.97	0.969	0.972	0.968	0.967	0.969
RESET test	0.223	0.257	0.157	0.217	0.083	0.199	0.294	0.217

Table A-30: WW2 with water resource driver

Model	WW2 Model 1	WW2 Model 18	WW2 Model 19	WW2 Model 20	WW2 Model 21	WW2 Model 22	WW2 Model 23	WW2 Model 24
Version	PR19	% of DI from pumped & impounding reservoirs	% of DI from pumped reservoirs	% of DI from impounding reservoirs	Number of pumped reservoirs	Number of impounding reservoirs	Number of reservoirs	% DI from all pumped sources
Properties (log)	1.059***	1.044***	1.063***	1.051***	1.051***	1.112***	1.040***	1.051***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
WAD - LAD (log)	-1.832***	-1.647***	-1.383***	-1.911***	-1.917***	-1.777***	-1.864***	-1.914***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Squared WAD - LAD (log)	0.128***	0.116***	0.097***	0.134***	0.135***	0.122***	0.131***	0.134***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Weighted average complexity (log)	0.430***	0.332**	0.316**	0.416***	0.411**	0.686***	0.639**	0.416***
	{0.001}	{0.019}	{0.032}	{0.002}	{0.012}	{0.000}	{0.018}	{0.002}
Boosters per length (log)	0.334**	0.377**	0.429**	0.323**	0.443**	0.485***	0.322**	0.322**
	{0.019}	{0.024}	{0.019}	{0.036}	{0.010}	{0.000}	{0.029}	{0.036}
DI from reservoirs (%)		0.002**						
		{0.011}						
DI from pumped reservoirs (%)			0.003**					
			{0.012}					
DI from impounding reservoirs (%)				0.001				
				{0.451}				
Nr. of pumped reservoirs (log)					0.023			
					{0.701}			
Nr. of impounding reservoirs (log)						-0.016*		
						{0.093}		
Nr. of reservoirs (log)							0.017	

							{0.471}	
DI from all pumped sources (%)								-0.001
								{0.430}
Constant	-2.589***	-2.856***	-3.730***	-2.277**	-1.749*	-3.163***	-2.669***	-2.157**
	{0.001}	{0.000}	{0.000}	{0.011}	{0.089}	{0.000}	{0.004}	{0.024}
R-squared	0.971	0.969	0.97	0.97	0.972	0.972	0.969	0.97
RESET test	0.122	0.213	0.141	0.132	0.048	0.088	0.168	0.132

CEPA economics matters

A.1.10. Time trend

Table A-31: Models with time trend

Model	WRP1 Model 1	WRP2 Model 1	TWD Model 1	WW1 Model 1	WW2 Model 1	WRP1 Model 34	WRP2 Model 30	TWD Model 17	WW1 Model 29	WW2 Model 25
Version	PR19	PR19	PR19	PR19	PR19	Time trend	Time trend	Time trend	Time trend	Time trend
Properties (log)	1.074***	1.069***		1.071***	1.059***	1.067***	1.066***		1.067***	1.059***
	{0.000}	{0.000}		{0.000}	{0.000}	{0.000}	{0.000}		{0.000}	{0.000}
Water treated in bands 3-6 (%)	0.006***			0.004***		0.004**			0.003***	
	{0.000}			{0.000}		{0.024}			{0.003}	
WAD - LAD (log)	-1.614***	-1.412***	-2.946***	-2.094***	-1.832***	-1.329**	-1.189*	-2.739***	-1.944***	-1.773***
	{0.000}	{0.005}	{0.000}	{0.000}	{0.000}	{0.011}	{0.066}	{0.000}	{0.000}	{0.000}
Squared WAD - LAD (log)	0.101***	0.087***	0.235***	0.147***	0.128***	0.080**	0.07	0.218***	0.136***	0.123***
	{0.000}	{0.009}	{0.000}	{0.000}	{0.000}	{0.021}	{0.102}	{0.000}	{0.000}	{0.000}
Weighted average complexity (log)		0.377			0.430***		0.164			0.328***
		{0.123}			{0.001}		{0.408}			{0.003}
Lengths of main (log)			1.077***					1.070***		
			{0.000}					{0.000}		
Boosters per length (log)			0.437***	0.335**	0.334**			0.409***	0.322**	0.321**
			{0.002}	{0.032}	{0.019}			{0.006}	{0.036}	{0.026}
year						0.011	0.012	0.011***	0.009	0.008
						{0.263}	{0.161}	{0.009}	{0.101}	{0.150}
Constant	-5.093***	-5.805***	4.723***	-1.565*	-2.589***	-27.688	-30.532*	-18.487**	-19.814*	-18.006*
	{0.000}	{0.000}	{0.002}	{0.074}	{0.001}	{0.166}	{0.085}	{0.028}	{0.062}	{0.080}
R-squared	0.917	0.907	0.957	0.97	0.971	0.912	0.901	0.958	0.969	0.97
RESET test	0.439	0.324	0.102	0.223	0.122	0.749	0.66	0.315	0.469	0.278

A.2. Wastewater models

A.2.1. Wastewater network plus models

Table A-32: WWWN+ baseline candidates

Driver	WWWN+ Model 1	WWWN+ Model 2	WWWN+ Model 3	WWWN+ Model 4	WWWN+ Model 5	WWWN+ Model 6	WWWN+ Model 7	WWWN+ Model 8	WWWN+ Model 9	WWWN+ Model 10
Load (log)	0.767*** {0.000}			0.846*** {0.000}	0.799*** {0.000}	0.620*** {0.000}	0.742*** {0.000}	0.720*** {0.000}	0.709*** {0.000}	0.763*** {0.000}
Sewer length (log)		0.919*** {0.000}								
Number of properties (log)			0.822*** {0.000}							
Load treated in size bands 1 to 3 (%)				0.024 {0.152}					0.035 {0.123}	0.023* {0.090}
Load treated in size band 6 (%)					-0.003 {0.549}					
Properties per sewer length (log)						0.872** {0.043}			0.328 {0.350}	
LAD Weighted average density (log)							0.049 {0.456}	-1.024 {0.432}		0.715 {0.268}
Square LAD Weighted average density (log)								0.074 {0.391}		-0.054 {0.219}
Pumping capacity per sewer length (log)									0.325*** {0.002}	0.368*** {0.000}
Load treated with ammonia consent ≤ 3mg/l									0.005*** {0.000}	0.006*** {0.000}
Constant	-4.240*** {0.000}	-4.461*** {0.000}	-6.471*** {0.000}	-5.329*** {0.000}	-4.437*** {0.000}	-5.616*** {0.000}	-4.271*** {0.000}	-0.111 {0.983}	-5.115*** {0.003}	-6.891** {0.012}
Observations	110	110	110	110	110	110	110	110	110	110
R2	0.911	0.846	0.907	0.913	0.91	0.902	0.908	0.902	0.948	0.959

RESET test	0.007	0.179	0.004	0.012	0.01	0.003	0.003	0.000	0.417	0.001

Driver	WWWN + Model 11	WWWN + Model 12	WWWN + Model 13	WWWN + Model 14	WWWN + Model 15	WWWN + Model 16	WWWN + Model 17	WWWN + Model 18	WWWN + Model 19	WWWN + Model 20	WWWN + Model 21
Load (log)	0.672***	0.707***					0.664***	0.697***	0.727***	0.691***	0.646***
	{0.000}	{0.000}					{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Sewer length (log)			0.758***	0.749***	0.702***	0.765***					
			{0.000}	{0.000}	{0.000}	{0.000}					
Number of properties (log)											
Load treated in size bands 1 to 3 (%)			0.035	-0.019*					0.023*		
			{0.252}	{0.051}					{0.073}		
Load treated in size band 6 (%)	-0.008	-0.002			-0.006	0.010***				-0.004	
	{0.411}	{0.653}			{0.518}	{0.001}				{0.167}	
Properties per sewer length (log)	0.352		1.031**		0.969		-0.104				
	{0.585}		{0.022}		{0.126}		{0.540}				
LAD Weighted average density (log)		0.112		-1.249		-1.188***		0.026			
		{0.839}		{0.130}		{0.008}		{0.967}			
Square LAD Weighted average density (log)		-0.015		0.079		0.068**		-0.01			
		{0.702}		{0.154}		{0.027}		{0.811}			
Pumping capacity per sewer length (log)	0.313***	0.360***	0.307**	0.455***	0.297**	0.487***	0.387***	0.361***	0.380***	0.370***	0.367***
	{0.004}	{0.001}	{0.022}	{0.001}	{0.037}	{0.000}	{0.000}	{0.000}	{0.000}	{0.001}	{0.000}
Load treated with ammonia consent ≤ 3mg/l	0.005***	0.006***	0.005***	0.006***	0.005***	0.006***	0.005***	0.006***	0.005***	0.005***	0.005***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Constant	-3.999**	-3.674*	-6.928***	2.01	-5.483***	1.076	-2.830***	-3.287	-4.106***	-3.228***	-2.984***
	{0.011}	{0.065}	{0.002}	{0.504}	{0.001}	{0.534}	{0.000}	{0.189}	{0.000}	{0.000}	{0.000}
Observations	110	110	110	110	110	110	110	110	110	110	110
R2	0.944	0.957	0.94	0.947	0.936	0.961	0.949	0.958	0.952	0.949	0.947
RESET test	0.689	0.000	0.528	0.000	0.81	0.000	0.672	0.565	0.478	0.677	0.572

A.2.2. Rainfall

Table A-33: WWWN+ Rainfall

Driver	Baseline (WWWN+ Model 21)	WWWN+ Model 22	WWWN+ Model 23
Load (log)	0.646***	0.587***	0.650***
	{0.000}	{0.000}	{0.000}
Pumping capacity per sewer length (log)	0.367***	0.371***	0.355***
	{0.000}	{0.000}	{0.000}
Load treated with ammonia consent ≤ 3mg/l	0.005***	0.005***	0.005***
	{0.000}	{0.000}	{0.000}
Urban rainfall (log)		0.089***	
		{0.003}	
Urban rainfall per sewer length (log)			0.076**
			{0.017}
Constant	-2.984***	-2.940***	-2.807***
	{0.000}	{0.000}	{0.000}
Observations	110	110	110
R2	0.947	0.956	0.953
RESET test	0.572	0.376	0.163

Table A-34: SWC Rainfall

Driver	Baseline (PR19 SWC1)	SWC1 Model 1	SWC1 Model 2	Baseline (PR19 SWC2)	SWC2 Model 1	SWC2 Model 2
Pumping capacity per sewer length (log)	0.344**	0.357**	0.357**	0.604***	0.568***	0.568***
	{0.012}	{0.017}	{0.017}	{0.000}	{0.000}	{0.000}
Sewer length (log)	0.804***	0.725***	0.842***	0.859***	0.713***	0.867***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Properties per sewer length (log)	1.043***	0.972***	0.972***			
	{0.000}	{0.000}	{0.000}			
LAD Weighted average density (log)				-2.480**	-2.102***	-2.102***
				{0.021}	{0.000}	{0.000}
Square LAD Weighted average density (log)				0.181***	0.158***	0.158***
				{0.010}	{0.000}	{0.000}
Urban rainfall		0.116***			0.155***	
		{0.000}			{0.000}	
Urban rainfall per sewer length (log)			0.116***			0.155***
			{0.000}			{0.000}
Constant	-7.956***	-7.760***	-7.760***	3.606	2.51	2.51
	{0.000}	{0.000}	{0.000}	{0.395}	{0.245}	{0.245}
Observations	110	110	110	110	110	110
R2	0.917	0.92	0.92	0.895	0.917	0.917
RESET test	0.356	0.17	0.17	0.269	0.670	0.670

A.2.3. Treatment complexity

Table A-35: SWT1 Treatment complexity

Driver	Baseline (PR19 SWT1)	SWT1 Model 1	SWT1 Model 2	SWT1 Model 3	SWT1 Model 4	SWT1 Model 5	SWT1 Model 6	SWT1 Model 7	SWT1 Model 8	SWT1 Model 9	SWT1 Model 10
Load (log)	0.653*** {0.000}	0.832*** {0.000}	0.864*** {0.000}	0.888*** {0.000}	0.777*** {0.000}	0.812*** {0.000}	0.818*** {0.000}	0.640*** {0.000}	0.849*** {0.000}	0.635*** {0.000}	0.642*** {0.000}
Load treated with ammonia consent ≤ 3mg/l (%)	0.006*** {0.000}							0.005*** {0.000}			
Load treated in size bands 1 to 3 (%)	0.029 {0.211}	0.032* {0.092}	0.033* {0.083}	0.034* {0.062}	0.03 {0.145}	0.029 {0.203}	0.025 {0.330}	0.027 {0.279}	0.034 {0.108}	0.029 {0.241}	0.029 {0.226}
Load treated with ammonia consent ≤ 1mg/l (%)		0.01 {0.172}									
Load treated with phosphorous consent ≤ 0.5mg/l (%)			0.015 {0.248}								
Load treated with phosphorous consent ≤ 1mg/l (%)				0.000 {0.891}							
Load treated with BOD consent ≤ 7mg/l (%)					0.032*** {0.000}			0.017** {0.036}	0.015 {0.818}		
Load treated with BOD consent ≤ 10mg/l (%)						0.005 {0.551}					
Load treated with UV ≤ 30mW/s/cm2 (%)							-0.002 {0.617}				
Sq. of load treated with BOD consent ≤ 7mg/l (%)									0.000 {0.933}		
NH3≤3 + BOD≤7 + P≤0.5 (%)										0.006*** {0.000}	
NH3≤3 + P≤0.5 (%)											0.006***

											{0.000}
Constant	-3.734***	-5.897***	-6.287***	-6.587***	-5.206***	-5.669***	-5.634**	-3.554**	-6.111***	-3.504**	-3.597***
	{0.004}	{0.000}	{0.000}	{0.000}	{0.001}	{0.001}	{0.013}	{0.013}	{0.000}	{0.013}	{0.007}
Observations	110	110	110	110	110	110	110	110	110	110	110
R2	0.854	0.874	0.868	0.866	0.841	0.862	0.872	0.842	0.863	0.851	0.855
RESET test	0.056	0.008	0.001	0.016	0.000	0.005	0.009	0.000	0.000	0.006	0.023

Table A-36: SWT2 Treatment complexity

Driver	Baseline (PR19 SWT2)	SWT2 Model 1	SWT2 Model 2	SWT2 Model 3	SWT2 Model 4	SWT2 Model 5	SWT2 Model 6	SWT2 Model 7	SWT2 Model 8	SWT2 Model 9	SWT2 Model 10
Load (log)	0.658*** {0.000}	0.820*** {0.000}	0.850*** {0.000}	0.864*** {0.000}	0.783*** {0.000}	0.778*** {0.000}	0.796*** {0.000}	0.659*** {0.000}	0.804*** {0.000}	0.646*** {0.000}	0.650*** {0.000}
Load treated in size band 6 (%)	-0.009* {0.097}	-0.008 {0.172}	-0.008 {0.164}	-0.008 {0.195}	-0.009 {0.102}	-0.007 {0.242}	-0.005 {0.523}	-0.010* {0.086}	-0.008 {0.157}	-0.010* {0.081}	-0.010* {0.087}
Load treated with ammonia consent ≤ 3mg/l (%)	0.006*** {0.000}							0.005*** {0.000}			
Load treated with ammonia consent ≤ 1mg/l (%)		0.01 {0.176}									
Load treated with phosphorous consent ≤ 0.5mg/l (%)			0.015 {0.237}								
Load treated with phosphorous consent ≤ 1mg/l (%)				0.000 {0.969}							
Load treated with BOD consent ≤ 7mg/l (%)					0.034*** {0.000}			0.018** {0.019}	0.022 {0.696}		
Load treated with BOD consent ≤ 10mg/l (%)						0.007 {0.357}					
Load treated with UV ≤ 30mW/s/cm2 (%)							-0.002 {0.667}				
Sq. of load treated with BOD consent ≤ 7mg/l (%)									0.000 {0.994}		
NH_3≤3 + BOD≤7 + P≤0.5 (%)										0.006*** {0.000}	
NH_3≤3 + P≤0.5 (%)											0.006***

											{0.000}
Constant	-2.965***	-4.981***	-5.347***	-5.564***	-4.444***	-4.579***	-4.882***	-2.925***	-4.782***	-2.783***	-2.849***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.005}	{0.002}	{0.000}	{0.001}	{0.000}
Observations	110	110	110	110	110	110	110	110	110	110	110
R2	0.855	0.871	0.86	0.859	0.841	0.857	0.867	0.846	0.853	0.853	0.855
RESET test	0.142	0.012	0.002	0.019	0.001	0.011	0.005	0.000	0.000	0.024	0.057

Table A-37: Network topography and coastal population

Driver	Baseline (PR19 SWC1)	SWC1 Model 3	Baseline (PR19 SWC2)	SWC2 Model 3
Sewer length (log)	0.804***	0.808***	0.859***	0.599**
	{0.000}	{0.000}	{0.000}	{0.031}
Pumping capacity per sewer length (log)	0.344**	0.317	0.604***	0.813**
	{0.012}	{0.220}	{0.000}	{0.023}
Properties per sewer length (log)	1.043***	1.054***		
	{0.000}	{0.000}		
Population in coastal areas (%)		0.000		-0.012
		{0.946}		{0.324}
LAD Weighted average density (log)			-2.480**	-2.832**
			{0.021}	{0.032}
Square LAD Weighted average density (log)			0.181***	0.206**
			{0.010}	{0.018}
Constant	-7.956***	-8.042***	3.606	7.82
	{0.000}	{0.000}	{0.395}	{0.240}
Observations	110	110	110	
R2	0.917	0.918	0.895	0.891
RESET test	0.356	0.345	0.269	0.193

Table A-38: Network topography and coastal population

Driver	Baseline (PR19 SWT1)	SWT1 Model 11	SWT1 Model 12	Baseline (PR19 SWT2)	SWT2 Model 11	SWT2 Model 12
Load (log)	0.653***	0.833***	0.748***	0.658***	0.890***	0.736***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Load treated with ammonia consent ≤ 3mg/l	0.006***	0.006***	0.005***	0.006***	0.006***	0.006***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Load treated in size bands 1 to 3 (%)	0.029	0.032*	0.034*			
	{0.211}	{0.066}	{0.064}			
Load treated in size band 6 (%)				-0.009*	-0.012***	-0.010**
				{0.097}	{0.002}	{0.043}
Population in coastal areas (%)		0.0092**			0.011***	
		{0.025}			{0.003}	
Pumping capacity per sewer length (log)			0.333*			0.322
			{0.077}			{0.113}
Constant	-3.734***	-6.198***	-5.082***	-2.965***	-5.931***	-4.051***
	{0.004}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Observations	110	110	110	110	110	110
R2	0.854	0.887	0.891	0.855	0.898	0.891
RESET test	0.056	0.000	0.025	0.142	0.045	0.178

Table A-39: Coastal WWWN+

Driver	Baseline (WWWN+ Model 21)	WWWN+ Model 24
Load (log)	0.647***	0.654***
	{0.000}	{0.000}
Pumping capacity per sewer length (log)	0.367***	0.353
	{0.000}	{0.180}
Load treated with ammonia consent ≤ 3mg/l	0.005***	0.005***
	{0.000}	{0.000}
Population in coastal areas (%)		0.000
		{0.943}
Constant	-2.995***	-3.086**
	{0.000}	{0.043}
Observations	110	110
R2	0.947	0.948
RESET test	0.569	0.243

A.2.4. Sewage treatment economies of scale

Table A-40: SWT Economies of scale

Driver	Baseline (PR19 SWT2)	SWT2 Model 13	SWT2 Model 14	SWT2 Model 15	SWT2 Model 16	SWT2 Model 17	SWT2 Model 18	SWT2 Model 19	SWT2 Model 20
Load (log)	0.658*** {0.000}	0.723*** {0.000}	0.669*** {0.000}	0.713*** {0.000}	0.686*** {0.000}	0.665*** {0.000}	0.747*** {0.000}	0.770*** {0.000}	0.788*** {0.000}
Load treated with ammonia consent ≤ 3mg/l	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}
Load treated in size band 6 (≥ 25,000 people) (%)	-0.009* {0.097}								
Load treated in STWs ≥ 100,000 people (%)		-0.008*** {0.007}							
Load treated in STWs ≥ 125,000 people (%)			-0.005*** {0.008}						
Load treated in STWs ≥ 250,000 people (%)				-0.007*** {0.003}					
Load treated in STWs ≥ 500,000 people (%)					-0.009*** {0.000}				
Load treated in STWs ≥ 1,000,000 people (%)						-0.007*** {0.002}			
Weighted average CEPA size band (log)							-2.015*** {0.000}		
Weighted average ANH size band (log)								-1.485*** {0.000}	
Weighted average treatment size (log)									-0.242*** {0.000}

Constant	-2.965***	-4.072***	-3.567***	-4.142***	-3.848***	-3.730***	-1.091	-2.292***	-3.001***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.221}	{0.001}	{0.000}
Observations	110	110	110	110	110	110	110	110	110
R2	0.855	0.869	0.87	0.899	0.919	0.879	0.887	0.899	0.911
RESET test	0.142	0.272	0.221	0.463	0.827	0.389	0.228	0.431	0.849

Table A-41: WWWN+ Economies of scale

Driver	Baseline (WWWN+ Model 21)	WWWN+ Model 19	WWWN+ Model 25	WWWN+ Model 26	WWWN+ Model 27	WWWN+ Model 28	WWWN+ Model 29	WWWN+ Model 30
Load (log)	0.646***	0.727***	0.691***	0.686***	0.706***	0.696***	0.676***	0.617***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Load treated with ammonia consent ≤ 3mg/l	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.006***	0.005***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Pumping capacity per sewer length (log)	0.367***	0.380***	0.370***	0.359***	0.347***	0.330***	0.259***	0.381***
	{0.000}	{0.000}	{0.001}	{0.000}	{0.000}	{0.002}	{0.003}	{0.000}
Load treated in size bands 1-3 (≤ 2,000 people)		0.023*						
		{0.073}						
Load treated in size band 6 (≥ 25,000 people) (%)			-0.004					
			{0.167}					
Load treated in STWs ≥ 100,000 people (%)				-0.002				
				{0.204}				
Load treated in STWs ≥ 125,000 people (%)					-0.003**			
					{0.021}			
Load treated in STWs ≥ 250,000 people (%)						-0.003		
						{0.162}		
Load treated in STWs ≥ 500,000 people (%)							-0.004***	
							{0.002}	
Load treated in STWs ≥ 1,000,000 people (%)								0.002
								{0.487}
Constant	-2.984***	-4.106***	-3.228***	-3.374***	-3.578***	-3.505***	-3.238***	-2.628***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Observations	110	110	110	110	110	110	110	110
R2	0.947	0.952	0.949	0.949	0.95	0.954	0.959	0.941
RESET test	0.572	0.478	0.677	0.700	0.783	0.798	0.918	0.176

Driver	Baseline (WWWN+ Model 21)	WWWN+ Model 31	WWWN+ Model 32	WWWN+ Model 33
Load (log)	0.646***	0.715***	0.703***	0.714***
	{0.000}	{0.000}	{0.000}	{0.000}
Load treated with ammonia consent ≤ 3mg/l	0.005***	0.005***	0.005***	0.005***
	{0.000}	{0.000}	{0.000}	{0.000}
Pumping capacity per sewer length (log)	0.367***	0.332***	0.334***	0.295***
	{0.000}	{0.001}	{0.001}	{0.002}
Weighted average CEPA size band (log)		-0.843***		
		{0.002}		
Weighted average ANH size band (log)			-0.464	
			{0.158}	
Weighted average treatment size (log)				-0.092**
				{0.012}
Constant	-2.984***	-2.283***	-2.806***	-2.929***
	{0.000}	{0.000}	{0.000}	{0.000}
Observations	110	110	110	110
R2	0.947	0.953	0.954	0.956
RESET test	0.572	0.801	0.771	0.901

A.2.5. Alternative density driver

Table A-42: SWC2 Density

Driver	Baseline (PR19 SWC2)	SWC2 Model 4	SWC2 Model 5	SWC2 Model 6	SWC2 Model 7	SWC2 Model 8
Sewer length (log)	0.859***	0.893***	0.888***	0.847***	0.861***	0.852***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Pumping capacity per sewer length (log)	0.604***	0.598***	0.586***	0.594***	0.542***	0.554***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.001}	{0.000}
LAD Weighted average density (log)	-2.480**	0.210**				

	{0.021}	{0.043}				
Square LAD Weighted average density (log)	0.181***					
	{0.010}					
LAD from MSOA Weighted average density (log)			0.212**	-2.291**		
			{0.022}	{0.041}		
Square LAD from MSOA Weighted average density (log)				0.169**		
				{0.021}		
MSOA Weighted average density (log)					0.354***	-5.051*
					{0.005}	{0.060}
Square MSOA Weighted average density (log)						0.336**
						{0.039}
Constant	3.606	-6.663***	-6.609***	3.016	-7.572***	14.241
	{0.395}	{0.000}	{0.000}	{0.501}	{0.000}	{0.195}
Observations	110	110	110	110	110	110
R2	0.895	0.884	0.889	0.897	0.889	0.895
RESET test	0.269	0.306	0.307	0.326	0.254	0.399

Table A-43: WWWN+ Density

Driver	Baseline (WWWN+ Model 21)+	WWWN+ Model 34	WWWN+ Model 35	WWWN+ Model 36	WWWN+ Model 37	WWWN+ Model 38	WWWN+ Model 39
Load (log)	0.646*** {0.000}	0.696*** {0.000}	0.697*** {0.000}	0.695*** {0.000}	0.697*** {0.000}	0.711*** {0.000}	0.710*** {0.000}
Pumping capacity per sewer length (log)	0.367*** {0.000}	0.364*** {0.000}	0.361*** {0.000}	0.368*** {0.000}	0.364*** {0.000}	0.399*** {0.000}	0.396*** {0.000}
Load treated with ammonia consent ≤ 3mg/l (%)	0.005*** {0.000}	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}
LAD Weighted average density (log)		-0.128*** {0.000}	0.026 {0.967}				
Square LAD Weighted average density (log)			-0.01 {0.811}				
LAD from MSOA Weighted average density (log)				-0.123*** {0.000}	0.047 {0.943}		
Square LAD from MSOA WAD density (log)					-0.011 {0.800}		
MSOA Weighted average density (log)						-0.224*** {0.000}	-0.323 {0.801}
Square MSOA Weighted average density (log)							0.007 {0.935}
Constant	-2.984*** {0.000}	-2.706*** {0.000}	-3.287 {0.189}	-2.732*** {0.000}	-3.378 {0.206}	-2.073*** {0.000}	-1.687 {0.746}
Observations	110	110	110	110	110	110	110
R2	0.947	0.958	0.958	0.957	0.957	0.961	0.961
RESET test	0.572	0.739	0.565	0.741	0.003	0.751	0.473

A.2.6. Network reinforcement drivers

Table A-44: SWC Network reinforcement

Driver	Baseline (PR19 SWC1)	SWC1 Model 4	SWC1 Model 5	SWC1 Model 6	Baseline (PR19 SWC2)	SWC2 Model 9	SWC2 Model 10	SWC2 Model 11

Sewer length (log)	0.804***	0.822***	0.816***	0.801***	0.859***	0.867***	0.865***	0.862***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Pumping capacity per sewer length (log)	0.344**	0.379**	0.383**	0.320***	0.604***	0.617***	0.609***	0.597***
	{0.012}	{0.013}	{0.017}	{0.010}	{0.000}	{0.000}	{0.000}	{0.000}
Properties per sewer length (log)	1.043***	0.989***	1.040***	1.042***				
	{0.000}	{0.000}	{0.000}	{0.000}				
LAD Weighted average density (log)					-2.480**	-2.407**	-2.270**	-2.430**
					{0.021}	{0.010}	{0.022}	{0.023}
Square LAD Weighted average density (log)					0.181***	0.176***	0.167***	0.178**
					{0.010}	{0.004}	{0.009}	{0.011}
New properties as a % of total properties (%)		-0.087*				-0.130***		
		{0.082}				{0.004}		
Property growth (%)			-0.063*				-0.072**	
			{0.066}				{0.027}	
Population growth (%)				-0.012*				-0.012*
				{0.082}				{0.073}
Constant	-7.956***	-7.908***	-8.047***	-7.905***	3.606	3.349	2.82	3.409
	{0.000}	{0.000}	{0.000}	{0.000}	{0.395}	{0.355}	{0.458}	{0.422}
Observations	110	110	110	110	110	110	110	110
R2	0.917	0.915	0.916	0.918	0.895	0.906	0.903	0.898
RESET test	0.356	0.395	0.004	0.503	0.269	0.112	0.091	0.001

Table A-45: WWWN+ Network reinforcement

Driver	Baseline (WWWN+ Model 21)+	WWWN+ Model 40	WWWN+ Model 41	WWWN+ Model 42
Load (log)	0.646***	0.630***	0.626***	0.657***
	{0.000}	{0.000}	{0.000}	{0.000}
Pumping capacity per sewer length (log)	0.367***	0.386***	0.389***	0.365***
	{0.000}	{0.000}	{0.000}	{0.000}

Load treated with ammonia consent ≤ 3mg/l (%)	0.005***	0.006***	0.006***	0.005***
	{0.000}	{0.000}	{0.000}	{0.000}
New properties as a % of total properties (%)		-0.087		
		{0.108}		
Property growth (%)			-0.062**	
			{0.017}	
Population growth (%)				-0.011
				{0.111}
Constant	-2.984***	-2.743***	-2.710***	-3.102***
	{0.000}	{0.000}	{0.000}	{0.000}
Observations	110	110	110	110
R2	0.947	0.946	0.947	0.947
RESET test	0.572	0.57	0.529	0.536

A.2.7. Time trend

Table A-46: Time trend SWC

Driver	Baseline (PR19 SWC1)	SWC1 Model 6	Baseline (PR19 SWC2)	SWC2 Model 12
Sewer length (log)	0.804***	0.804***	0.859***	0.854***
	{0.000}	{0.000}	{0.000}	{0.000}
Pumping capacity per sewer length (log)	0.344**	0.344**	0.604***	0.564***
	{0.012}	{0.011}	{0.000}	{0.000}
Properties per sewer length (log)	1.043***	1.041***		
	{0.000}	{0.000}		
LAD Weighted average density (log)			-2.480**	-2.536**
			{0.021}	{0.016}
Square LAD Weighted average density (log)			0.181***	0.184***
			{0.010}	{0.007}
Year		0.000		0.003
		{0.985}		{0.523}
Constant	-7.956***	-8.155	3.606	-3.019
	{0.000}	{0.467}	{0.395}	{0.783}
Observations	110	110	110	110
R2	0.917	0.917	0.895	0.897
RESET test	0.356	0.348	0.269	0.469

Table A-47: Time trend SWT

Driver	Baseline (PR19 SWT1)	SWT1 Model 13	Baseline (PR19 SWT2)	SWT2 Model 21
Load (log)	0.653***	0.728***	0.658***	0.740***
	{0.000}	{0.000}	{0.000}	{0.000}
Load treated with ammonia consent < 3mg/l	0.006***	0.004***	0.006***	0.004***
	{0.000}	{0.003}	{0.000}	{0.003}

Load treated in size bands 1 to 3 (%)	0.029	0.039		
	{0.211}	{0.189}		
Load treated in size band 6 (%)			-0.009*	-0.013
			{0.097}	{0.101}
Year		0.012*		0.013*
		{0.068}		{0.064}
Constant	-3.734***	-29.842**	-2.965***	-29.641**
	{0.004}	{0.034}	{0.000}	{0.038}
Observations	110	110	110	110
R2	0.854	0.868	0.855	0.861
RESET test	0.056	0.021	0.142	0.034

Table A-48: Time trend WWWN+

Driver	Baseline (WWWN+ Model 21)+	WWWN+ Model 43
Load (log)	0.646***	0.655***
	{0.000}	{0.000}
Pumping capacity per sewer length (log)	0.367***	0.340***
	{0.000}	{0.000}
Load treated with ammonia consent < 3mg/l	0.005***	0.005***
	{0.000}	{0.000}
Year		0.003
		{0.407}
Constant	-2.984***	-9.522
	{0.000}	{0.235}
Observations	110	110
R2	0.947	0.948
RESET test	0.572	0.536

A.3. Bioresources models

A.3.1. Density

Table A-49: Bioresources 1 with alternative density drivers

Driver	BR1 Model 1	BR1 Model 2	BR1 Model 3	BR1 Model 4	BR1 Model 5	BR1 Model 6	BR1 Model 7	BR1 Model 8
Sludge produced (log)	1.172***	1.132***	1.176***	1.136***	1.119***	1.094***	1.119***	1.133***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
WAD - LAD (log)	-0.133				-2.247			
	{0.267}				{0.134}			
Load treated in STWs bands 1-3 (%)	0.063**	0.064**	0.063**	0.046**	0.050**	0.043	0.050**	0.042
	{0.011}	{0.016}	{0.011}	{0.045}	{0.034}	{0.117}	{0.040}	{0.114}
Squared WAD - LAD (log)					0.142			
					{0.160}			
WAD - MSOA (log)		-0.093				-4.223		
		{0.642}				{0.236}		
Squared WAD - MSOA (log)						0.252		
						{0.252}		
WAD - LAD from MSOA (log)			-0.139				-2.159	
			{0.217}				{0.145}	
Squared WAD - LAD from MSOA (log)							0.136	
							{0.176}	
Properties per sewer length (log)				-0.638				0.565
				{0.218}				{0.962}
Squared properties per sewer length (log)								-0.17
								{0.914}
Constant	-0.912	-0.946	-0.889	0.744	7.227	16.196	6.898	-1.346
	{0.310}	{0.479}	{0.312}	{0.613}	{0.216}	{0.270}	{0.234}	{0.952}
R-squared	0.82	0.815	0.821	0.812	0.828	0.82	0.829	0.809
RESET test	0.528	0.409	0.488	0.009	0.74	0.147	0.45	0.023

A.3.2. Economies of Scale

Table A-50: Bioresources 1 with alternative Economies of Scale drivers

Driver	BR1 Model 1	BR1 Model 9	BR1 Model 10	BR1 Model 11	BR1 Model 12	BR1 Model 13	BR1 Model 14	BR1 Model 15	BR1 Model 16
Sludge produced (log)	1.172***	1.092***	1.138***	1.157***	1.047***	0.977***	1.121***	1.123***	1.042***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
WAD - LAD (log)	-0.133	-0.201	-0.171	-0.07	-0.106	-0.463***	-0.109	-0.096	-0.224
	{0.267}	{0.213}	{0.291}	{0.668}	{0.575}	{0.004}	{0.512}	{0.576}	{0.246}
Load treated in STWs bands 1-3 (%)	0.063**								
	{0.011}								
Percentage of load treated in STWs > 100k people (%)		-0.004							
		{0.322}							
Percentage of load treated in STWs > 125k people (%)			-0.007						
			{0.244}						
Percentage of load treated in STWs > 250k people (%)				-0.010**					
				{0.029}					
Percentage of load treated in STWs > 500k people (%)					-0.006				
					{0.327}				
Percentage of load treated in STWs > 1m people (%)						0.012***			
						{0.000}			
Weighted average treatment band (CEPA) (%)							-1.851*		
							{0.064}		
Weighted average treatment band (ANH) (%)								-1.242**	
								{0.039}	
Weighted average treatment size (log)									-0.009

									{0.921}
Constant	-0.912	0.424	0.115	-0.708	-0.13	2.547***	2.826*	1.658	0.696
	{0.310}	{0.549}	{0.893}	{0.455}	{0.908}	{0.004}	{0.052}	{0.114}	{0.427}
R-squared	0.82	0.785	0.783	0.783	0.765	0.798	0.797	0.791	0.777
RESET test	0.528	0.046	0.425	0.326	0.328	0.124	0.292	0.342	0.241

Table A-51: Bioresources 2 with alternative Economies of Scale drivers

Driver	BR2 Model 1	BR2 Model 2	BR2 Model 3	BR2 Model 4	BR2 Model 5	BR2 Model 6	BR2 Model 7	BR2 Model 8	BR2 Model 9
Sludge produced (log)	1.134***	1.003***	1.064***	1.123***	0.997***	0.825***	1.080***	1.092***	0.964***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
STWs per property (log)	0.275								
	{0.174}								
Percentage of load treated in STWs > 100k people (%)		-0.006							
		{0.267}							
Percentage of load treated in STWs > 125k people (%)			-0.009						
			{0.189}						
Percentage of load treated in STWs > 250k people (%)				-0.011**					
				{0.024}					
Percentage of load treated in STWs > 500k people (%)					-0.007				
					{0.232}				
Percentage of load treated in STWs > 1m people (%)						0.005			
						{0.266}			
Weighted average treatment band (CEPA) (%)							-2.152**		
							{0.027}		
Weighted average treatment band (ANH) (%)								-1.448**	
								{0.019}	
Weighted average treatment size (log)									-0.076
									{0.393}
Constant	0.808	-0.504	-0.686	-1.034*	-0.631	-0.003	2.792**	1.507*	0.101
	{0.316}	{0.375}	{0.289}	{0.075}	{0.349}	{0.996}	{0.034}	{0.071}	{0.879}

R-squared	0.784	0.776	0.775	0.781	0.759	0.76	0.795	0.789	0.774
RESET test	0.374	0.714	0.513	0.32	0.139	0.47	0.442	0.494	0.687

A.3.3. Treatment complexity

Table A-52: Bioresources 1 with alternative treatment complexity drivers

Driver	BR1 Model 1	BR1 Model 17	BR1 Model 18	BR1 Model 19	BR1 Model 20	BR1 Model 21	BR1 Model 22	BR1 Model 23	BR1 Model 24
Sludge produced (log)	1.172***	1.187***	1.175***	1.174***	1.185***	1.216***	1.192***	1.218***	1.180***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
WAD - LAD (log)	-0.133	-0.117	-0.12	-0.127	-0.126	-0.108	-0.111	-0.116	-0.113
	{0.267}	{0.341}	{0.368}	{0.282}	{0.267}	{0.319}	{0.368}	{0.272}	{0.381}
Load treated in STWs bands 1-3 (%)	0.063**	0.061**	0.061**	0.064**	0.068***	0.073***	0.061**	0.074***	0.063**
	{0.011}	{0.025}	{0.022}	{0.011}	{0.010}	{0.005}	{0.025}	{0.003}	{0.016}
Load with ammonia below 1mg/l (%)			-0.004						
			{0.534}						
Load with phosphorus below 0.5mg/l (%)				-0.007					
				{0.454}					
Load with phosphorus below 1mg/l (%)					0.000				
					{0.923}				
NH below 3 + P below 1 (%)						-0.001			
						{0.592}			
NH below 3 + P below 0.5 (%)							-0.001		
							{0.139}		
NH below 1 + P below 1 (%)								-0.001	
								{0.650}	
NH below 1 + P below 0.5 (%)									-0.005
									{0.464}
Constant	-0.912	-1.066	-1.001	-0.966	-1.034	-1.311	-1.132	-1.274	-1.08
	{0.310}	{0.277}	{0.345}	{0.279}	{0.291}	{0.236}	{0.250}	{0.211}	{0.296}
R-squared	0.82	0.818	0.82	0.821	0.821	0.822	0.819	0.824	0.822
RESET test	0.528	0.801	0.427	0.121	0.602	0.688	0.828	0.574	0.436

Table A-53: Bioresources 2 with alternative treatment complexity drivers

Driver	BR2 Model 1	BR2 Model 10	BR2 Model 11	BR2 Model 12	BR2 Model 13	BR2 Model 14	BR2 Model 15	BR2 Model 16	BR2 Model 17
Sludge produced (log)	1.134*** {0.000}	1.149*** {0.000}	1.141*** {0.000}	1.133*** {0.000}	1.115*** {0.000}	1.134*** {0.000}	1.153*** {0.000}	1.125*** {0.000}	1.140*** {0.000}
STWs per property (log)	0.275 {0.174}	0.266 {0.198}	0.266 {0.226}	0.27 {0.180}	0.272 {0.190}	0.28 {0.179}	0.262 {0.208}	0.277 {0.181}	0.261 {0.231}
Load with ammonia below 1mg/l (%)			-0.003 {0.675}						
Load with phosphorus below 0.5mg/l (%)				-0.006 {0.505}					
Load with phosphorus below 1mg/l (%)					0.001 {0.559}				
NH below 3 + P below 1 (%)						0.000 {0.893}			
NH below 3 + P below 0.5 (%)							-0.001 {0.363}		
NH below 1 + P below 1 (%)								0.001 {0.773}	
NH below 1 + P below 0.5 (%)									-0.004 {0.587}
Constant	0.808 {0.316}	0.676 {0.439}	0.711 {0.476}	0.769 {0.351}	0.851 {0.306}	0.846 {0.356}	0.631 {0.484}	0.851 {0.330}	0.674 {0.497}
R-squared	0.784	0.783	0.784	0.786	0.782	0.784	0.783	0.783	0.785
RESET test	0.374	0.617	0.57	0.045	0.395	0.049	0.648	0.427	0.506

A.3.4. Time trend

Table A-54: Bioresources 1 with time trend

Driver	BR2 Model 1	BR1 Model 25
Sludge produced (log)	1.172***	1.163***
	{0.000}	{0.000}
WAD - LAD (log)	-0.133	-0.129
	{0.267}	{0.308}
Load treated in STWs bands 1-3 (%)	0.063**	0.061***
	{0.011}	{0.008}
year		-0.004
		{0.788}
Constant	-0.912	6.659
	{0.310}	{0.809}
R-squared	0.82	0.819
RESET test	0.528	0.525

Table A-55: Bioresources 2 with time trend

Driver	BR2 Model 1	BR2 Model 18
Sludge produced (log)	1.134***	1.122***
	{0.000}	{0.000}
STWs per property (log)	0.275	0.26
	{0.174}	{0.184}
year		-0.004
		{0.774}
Constant	0.808	8.696
	{0.316}	{0.752}
R-squared	0.784	0.784
RESET test	0.374	0.36

CEPA economics matters

Appendix B EFFICIENCY SCORES OF FINAL MODELS

B.1. WATER MODELS

B.1.1. Water resources plus models

Table B-1: Efficiency scores for final water resources plus models

Company	Baseline (WRP1 PR19)	Baseline (WRP2 PR19)	WRP1 Model 3 *MSOA instead of LAD density*	WRP1 Model 4 *LAD from MSOA density*	WRP1 Model 5 *Prop. per length instead of LAD density*	WRP2 Model 3 *MSOA instead of LAD density*	WRP2 Model 4 *LAD from MSOA density*	WRP2 Model 5 *Prop. per length instead of LAD density*
AFW	0.82	0.83	0.81	0.83	0.82	0.82	0.83	0.83
ANH	0.79	0.75	0.80	0.74	0.79	0.76	0.70	0.75
BRL	1.16	1.15	1.15	1.17	1.10	1.15	1.17	1.09
HDD	0.99	1.05	1.10	1.09	0.94	1.12	1.12	0.96
NES	1.06	1.08	1.08	1.07	1.08	1.10	1.10	1.10
NWT	1.18	1.19	1.19	1.18	1.21	1.20	1.19	1.22
PRT	0.72	0.68	0.66	0.71	0.66	0.64	0.68	0.63
SES	1.67	1.76	1.59	1.68	1.54	1.67	1.77	1.61
SEW	0.98	0.98	1.04	0.96	1.03	1.04	0.96	1.03
SRN	2.02	1.98	2.00	2.02	1.98	1.98	1.99	1.95
SSC	0.53	0.50	0.49	0.53	0.51	0.47	0.50	0.48
SVE	1.06	1.09	1.06	1.06	1.08	1.09	1.09	1.11
SWB	1.15	1.16	1.09	1.13	0.99	1.11	1.15	1.00
TMS	1.05	1.06	0.92	1.04	1.09	0.90	1.03	1.08
WSH	1.11	1.12	1.13	1.13	1.04	1.13	1.14	1.04
WSX	1.26	1.17	1.71	1.43	1.53	1.59	1.32	1.42
YKY	1.02	1.04	1.08	1.02	1.08	1.11	1.05	1.11

B.1.2. Treated water distribution models

Table B-2: Efficiency scores for final treated water distribution models

Company	Baseline (PR19 TWD)	TWD APH no booster, MSOA	TWD APH + booster, MSOA	TWD APH no booster, LAD from MSOA	TWD APH + booster, LAD from MSOA	TWD APH no booster, prop/length	TWD APH no booster, prop/length
AFW	1.12	1.32	1.24	1.29	1.19	1.19	1.08
ANH	1.38	1.19	1.35	1.18	1.38	1.18	1.38
BRL	1.33	1.13	1.07	1.29	1.22	1.28	1.21
HDD	0.96	1.15	1.02	1.11	0.96	1.05	0.99
NES	1.05	1.07	1.11	1.07	1.11	1.07	1.11
NWT	0.89	1.01	1.03	0.97	0.99	0.98	1.00
PRT	0.82	0.98	0.98	1.02	1.01	0.91	0.93
SES	1.15	0.98	1.11	0.96	1.09	0.97	1.14
SEW	1.28	1.01	1.02	1.10	1.15	1.05	1.04
SRN	0.96	1.00	0.91	1.13	1.02	1.09	0.97
SSC	1.20	1.10	1.15	1.01	1.03	0.94	0.95
SVE	1.01	1.04	1.03	0.97	0.95	1.04	1.00
SWB	0.77	0.71	0.75	0.72	0.75	0.75	0.82
TMS	1.11	0.97	1.01	1.02	1.06	1.01	1.07
WSH	1.14	1.26	1.20	1.26	1.18	1.23	1.19
WSX	1.01	1.11	1.02	1.08	1.04	1.13	1.06
YKY	1.25	1.21	1.16	1.32	1.28	1.21	1.14

B.1.3. Wholesale water models

Table B-3: Efficiency scores for final wholesale water models

Company	Baseline (PR19 WW1)	Baseline (PR19 WW2)	WW1 Length of mains, LAD from MSOA, boosters	WW2 Length of mains, LAD from MSOA, boosters	WW1 Length of mains, Lad from MSOA, APH, boosters	WW2 Length of mains, Lad from MSOA, APH, boosters	WW1 Length of mains, LAD from MSOA, APH, no boosters	WW2 Length of mains, LAD from MSOA, APH, no boosters
AFW	0.93	0.93	1.03	1.02	1.02	1.01	1.01	1.01
ANH	1.15	1.09	1.03	1.00	0.98	0.95	1.02	0.98
BRL	1.18	1.13	1.12	1.10	1.17	1.13	1.13	1.10
HDD	1.00	1.02	1.19	1.19	1.18	1.19	0.97	0.98
NES	1.01	1.02	1.09	1.09	1.09	1.09	1.07	1.07
NWT	0.99	1.0	1.12	1.12	1.10	1.10	1.11	1.11
PRT	0.78	0.77	0.90	0.89	0.94	0.93	0.83	0.83
SES	1.39	1.41	1.14	1.15	1.17	1.19	1.12	1.15
SEW	1.07	1.06	0.96	0.96	0.93	0.92	1.00	0.99
SRN	1.30	1.23	1.44	1.43	1.49	1.46	1.45	1.43
SSC	0.90	0.86	0.72	0.700	0.75	0.72	0.75	0.73
SVE	1.01	1.05	0.99	1.01	0.97	1.00	1.03	1.05
SWB	1.00	1.00	0.89	0.88	0.93	0.93	0.83	0.83
TMS	1.11	1.13	0.94	0.95	1.02	1.03	1.06	1.07
WSH	1.14	1.12	1.25	1.23	1.28	1.26	1.14	1.12
WSX	1.13	1.12	1.41	1.40	1.21	1.19	1.30	1.28
YKY	1.04	1.06	1.14	1.15	1.11	1.13	1.14	1.15

B.2. Wastewater models

B.2.1. Sewage collection models

Table B-4: Efficiency scores for final sewage collection models

Company	Baseline (PR19 SWC1)	SWC1 Model 2 *Urban rainfall*	Baseline (PR19 SWC2)	SWC2 Model 6 *MSOA*	SWC2 Model 8 *LAD from MSOA*	SWC2 Model 12 *Urban rainfall + MSOA*	SWC2 Model 13 *Urban rainfall + LAD from MSOA*
ANH	0.95	0.98	0.88	0.89	0.90	0.96	0.97
NES	0.92	0.96	1.05	1.04	1.04	1.06	1.06
NWT	1.06	1.00	1.11	1.11	1.10	1.02	1.02
SRN	0.95	0.94	1.00	1.00	1.00	0.99	1.00
SVH	0.94	0.95	1.05	1.04	1.07	1.07	1.03
SWB	1.11	1.13	0.96	0.96	0.98	1.01	0.99
TMS	1.13	1.17	0.96	0.96	0.94	0.97	0.99
WSH	1.04	0.96	1.04	1.06	1.09	1.00	0.98
WSX	0.91	0.93	0.87	0.88	0.86	0.88	0.90
YKY	1.03	1.02	1.21	1.21	1.16	1.13	1.18

B.2.2. Sewage treatment models

Table B-5: Efficiency scores for final sewage treatment models

Company	Baseline (PR19 SWT1)	Baseline (PR19 SWT2)	SWT2 Model 13 *Load treated ≥ 100,000 people*	SWT2 Model 18 *CEPA WAB*	SWT2 Model 20 *WATS*
ANH	1.12	1.03	1.04	0.97	0.93
NES	0.94	0.97	1.03	1.07	1.08
NWT	1.14	1.17	1.12	1.13	1.06
SRN	1.50	1.51	1.12	1.15	1.12
SVH	0.92	0.92	0.95	0.96	0.97
SWB	0.94	0.97	0.98	0.94	0.93
TMS	0.82	0.84	0.88	0.90	0.96
WSH	1.07	1.10	1.16	1.13	1.20
WSX	0.96	0.92	0.95	1.00	1.04
YKY	1.09	1.05	1.12	1.12	1.06

B.2.3. Wastewater network plus models

Table B-6: Efficiency scores for final wastewater network plus models

Company	Baseline (WWWN+ model 21)	WWWN+ Model 23 *Urban rainfall*	WWWN+ Model 31 *CEPA WAB*	WWWN+ Model 33 *WATS*	WWWN+ Model 44 *Urban rainfall + CEPA WAB*	WWWN+ Model 45 *Urban rainfall + CEPA WATS*
ANH	1.07	1.10	0.98	0.97	1.01	1.01
NES	0.95	0.97	1.04	1.03	1.06	1.06
NWT	1.06	1.03	1.06	1.03	1.02	0.99
SRN	1.05	1.04	1.06	1.05	1.06	1.05
SVH	0.99	0.99	0.96	0.97	0.97	0.98
SWB	1.04	1.05	0.99	0.99	1.00	1.01
TMS	0.92	0.93	0.95	0.97	0.97	1.00
WSH	1.05	1.00	1.07	1.09	1.01	1.03
WSX	0.93	0.95	0.94	0.95	0.96	0.97
YKY	1.03	1.02	1.05	1.03	1.04	1.01

B.3. Bioresources models

Table B-7: Efficiency scores for final bioresources models

Company	LAD from MSOA	No density variable	PR19 BR2	CEPA weighted average band
ANH	1.05	1.14	1.14	1.07
NES	0.68	0.68	0.60	0.67
NWT	0.87	0.87	0.84	0.83
SRN	0.96	0.96	0.98	0.83
SVH	0.89	0.89	0.83	0.91
SWB	1.00	0.93	1.11	1.01
TMS	1.02	0.93	1.11	1.06
WSH	1.44	1.53	1.47	1.72
WSX	1.18	1.15	1.16	1.09
YKY	1.33	1.40	1.28	1.36

UK

Queens House

55-56 Lincoln's Inn Fields

London WC2A 3LJ

T. **+44 (0)20 7269 0210**

E. **info@cepa.co.uk**

www.cepa.co.uk

cepa-ltd @cepaltd

Australia

Level 20, Tower 2 Darling Park

201 Sussex Street

Sydney NSW 2000

T. **+61 2 9006 1308**

E. **info@cepa.net.au**

www.cepa.net.au