Ofwat’s consultation on assessing base costs at PR24

Severn Trent response

WONDERFUL ON TAP

Executive Summary

We welcome the opportunity to engage with and are happy to support the high-level aspirations of Ofwat's emerging approach to cost assessment. We consider that cost assessment was largely effective at PR19, a point the CMA made in its final decision. Therefore, we believe that it is appropriate to seek evolutions to the PR19 approach rather than undertaking a dramatic 'reboot'. Given that context, we make the following recommendations.

- **Base model scope:** We believe the scope of base models should focus as closely as possible on base opex and capital maintenance (Botex) – with enhancement cost drivers subject to alternative assessments. The assessment of growth costs still needs improving. Further consideration of alternative approaches for assessing growth outside of econometric models is needed or increasing the explanatory power within econometric models for the subset of growth costs that would be most suited to inclusion in them.
- **Base model specifications**: We agree with using the PR19 models as the starting point for PR24. However, we identify several key areas where improvements should be considered, including the use of composite scale variables (to address the limitations of using specific measures) and the inclusion of a time trend (to reflect the changing circumstances being faced over time, particularly given intensifying climate change related pressures).
- **Adjustments:** We welcome Ofwat considering in more depth the basis of cost adjustments and their assessment. We consider symmetrical adjustments can provide a valuable way of managing model misspecifications where remedies within models are impractical or undesirable. However, there is also a need to consider the materiality of symmetrical adjustments and acknowledge that non-symmetrical adjustments will be needed where the costs are a function of specific circumstances or are not sufficiently captured in the historical cost dataset.
- **Future capital maintenance / asset health**: We consider that there is a material risk that future asset health pressures are not adequately accounted for in the current cost assessment approach. Consequently, it is appropriate to seek greater understanding on asset health. We suggest that using forecast rather than historical efficiency ratios when setting the catch-up efficiency challenge could be an appropriate improvement that should be considered further, but it would need to be carefully managed for it to deliver as intended. The inclusion of asset health metrics in models, coupled with appropriate mitigations to guard against endogeneity concerns should, in our view, not be discounted at this stage.
- **Cost service relationship:** We consider that further work is needed in this area otherwise it may lead to some undesirable outcomes. It is important that the underpinning definitions of capital maintenance and enhancement expenditure that flow into the cost models are recognised. Base expenditure should relate to the performance of the efficiency benchmark companies, and with an appropriate level of tolerance, this tends towards sector average performance levels (suggesting that performance is also impacted by external factors not accounted for in models). In line with the underlying definitions of base and enhancement, we think that the identified average level of performance should be projected forward.

We appreciate how comprehensive the consultation is and this has helpfully revealed areas that will require more consideration and testing. We remain keen to play our part in helping with improvements in these areas. In doing so we consider it is sensible to keep in mind the need for pragmatic solutions in terms of their potential to deliver robust models at PR24, given the likely future state of data, and the regulatory burden associated with some of the proposed analysis.

Contents

1. Overarching comments 4

2. Principles of PR24 base cost assessment 9

2.1 Using engineering logic to underpin the construction of econometric models 11

2.2 Using statistical methods and information to improve/validate models that have been identified based on engineering logic 13

3. Approach to wholesale base cost modelling at PR24 16

3.1 Scope of wholesale modelled base costs 16

3.2 Wholesale base cost modelling suite 22

3.3 Wholesale base cost drivers and explanatory variables 23

3.4 Sample period selection 42

4. Cost adjustment claims 46

5. Capital Maintenance and Asset Health 52

5.1 Potential approaches to improve consideration of future challenges 53

5.2 Asset Health Measures 56

6. Cost-service relationship 58

7. Residential retail cost assessment 62

Appendix A: Supporting evidence on wholesale base cost drivers and explanatory variables 63

1. Overarching comments

The key objective for Ofwat's base cost assessment approach should be to determine efficient base cost allowances so that customers only pay for the efficient cost of the services companies provide and companies have sufficient funding to provide those services. Ofwat's approach should also minimise risk for both customers and companies whilst ensuring sufficient opportunity to address future challenges in line with societal and customer expectations. We are broadly supportive of the emerging cost assessment approach Ofwat outlines for PR24 in its consultation document. However, we have identified the potential for improvements in some areas.

To help Ofwat navigate our response, we are making the following overarching comments relating to the specific aspects discussed in the 'Assessing base costs at PR24' consultation document. We then discuss them in more detail in the following chapters.

Principles of cost assessment (Chapter 2)

In our view, engineering logic should be the primary basis for specifying models and should ideally be supported by strong empirical evidence. We have started the process of improving this understanding of the engineering logic in five key areas in conjunction with external specialists. We consider we should avoid the temptation to select or develop models based on their statistical performance alone.

When developing and modifying models, fundamental consideration should be given to the quality of data that the sector is able to provide. If models are built using inconsistent or imprecise data, their corresponding robustness will be materially affected.

Once engineering logic and an appropriate level of data quality have been accepted, statistical tests can play a key role in refining models. This will improve statistical validity and can be used to distinguish between two similar models or competing cost drivers that are both supported by engineering logic. We have followed this logic throughout the analysis presented in our response.

We stress that models need to work in their wider cost assessment context. This includes:

- making sure that efficiency benchmarks are set on a sound basis that reflects the risk and uncertainty of models;
- enhancement costs are considered in a discrete but complementary way (enhancement cost assessment is not in the scope of this consultation, but we consider it is critical that clarity of the base and enhancement definitions are maintained);
- an appropriate cost adjustment process is in place so that model specification issues or weaknesses can be appropriately addressed (see chapter 4 for further discussion); and
- future challenges that the sector faces, which are unlikely to be adequately in historical data, are reflected in the wider cost assessment process (see chapter 5 for further discussion).

Base cost modelling at PR24 (Chapter 3)

The key objective when developing and using base econometric models is to make sure that they are fit for purpose and minimise the risk of random error disadvantaging customers or companies .

We believe the broad approach outlined in the consultation document is sensible and should not deviate substantially from the approach used at PR19. However, we have identified several areas where changes could be made to improve model robustness and coherence.

In short, we would like models to be developed in such a way that they best reveal efficiency, which is their main purpose, rather than uncertainty.

Scope of models (Chapter 3.1)

Fundamentally, econometrics are very effective for modelling some costs, and less so for others. We agree with the use of econometrics where there is a sound basis to support its use. However, where this is not the case and econometric modelling will increase uncertainty, we should seek and use alternative cost assessment processes.

The scope of econometric models should reflect as closely as possible opex and capital maintenance (together comprising base totex or "Botex") only. These costs are repeatable, incurred by all companies and therefore likely to be modellable. We do not think that growth costs should be included in base models as there is very little explanatory power for them in the current explanatory variables. Potential solutions are either to seek alternatives to model growth costs separately, or to introduce cost drivers that are sensitive to growth costs where they exist. In some areas (e.g. network reinforcement), this is not likely to be possible as the costs are highly related to local specific circumstances, rather than any company-wide factors.

Wholesale base cost modelling suite (Chapter 3.2)

We support the proposed modelling suite and believe it is logical to look at different levels of granularity. This allows for different cost drivers that may not be appropriate at other levels of specification to be included. It also allows for trade-offs to be made and/or differences in cost allocations to be mitigated. Consequently, it is sensible to explore how best to develop a wastewater network plus model.

It would be desirable to model water resources costs, but we acknowledge the challenges faced. As a minimum, more explanatory factors for water resources costs should be included in higher level models that include water resources expenditure.

Wholesale base cost drivers and explanatory variables (Chapter 3.3)

We have identified several variables that we believe were either mis-specified or didn't serve their intended purpose in the PR19 suite of models. Our suggested modifications here attempt to improve model robustness and predictive capability.

Composite scale variables

The use of composite scale variables can provide a richer view of how fundamental scale drivers relate to cost. We consider it is too simplistic to assume that scale effects can be fully accounted for by one variable. Principal Component Analysis (PCA) provides an appropriate way of aggregating scale drivers without the need for judgement to weight different variables. We note that Ofgem developed and used a composite scale variable for its RIIO-GD2 econometric modelling.

Density

The squared density term used at PR19 sought to explain the non-linear relationship between cost and density (i.e. that costs increase at both extremes of density). However, the existing density variables are not reflecting cost drivers as anticipated. Weighted average local authority population data suffers from specification issues and poorly reflects cost drivers that are focused on the extremes of density.

Therefore, we consider the squared density term should be removed. However, this then leaves no explanatory power linked to the extremes of density (which has strong engineering logic). Unfortunately, using high density and sparsity thresholds with current data is problematic. This is because high density is colinear with weighted average density and the current sparsity values poorly expose areas of extreme sparsity.

We have yet to find a perfect solution for density/sparsity related costs, but consider that the following areas merit further consideration:

- Our analysis at postcode sector level is promising and provides a much richer picture. We could seek ways of improving granularity of data and then include relevant variables in models.
- We could assume just a central view of density in models (weighted average, but not weighted average squared) then use symmetrical adjustments for sparsity and density (potentially using kurtosis, i.e. the 'tailedness' of the data).

Water resource cost drivers

Currently, no water resources cost drivers are used. Therefore, the water resources costs present in the water resources plus (WRP) models are just contributing to noise. We acknowledge the difficulties in finding appropriate drivers here, but there is a need to explore further and ensure that the underlying engineering logic is clear and evidenced.

Network complexity

Engineering logic points to the inclusion of both booster pumping stations per length and average pumping head (APH) in the treated water distribution (TWD) and wholesale water (WW) specifications. Therefore, in accordance with principle 1 (primacy of engineering logic), we advocate including both in models and not choosing between the two. The statistical evidence also shows that including both helps the models.

Water treatment complexity

There is potential for misspecification when using the current water treatment complexity data because:

- certain treatment processes may be in the wrong band;
- multiple 'higher cost' treatment processes might not be significantly more expensive than a single 'higher cost' treatment process; and
- single 'higher cost' treatment processes might not be significantly more expensive than multiple 'lower cost' treatment processes.

We don't yet have a robust solution to these issues but we have commissioned consultancy support to study the engineering logic and plan to share the results.

Accounting for changes over time – indicators and trends

The inclusion of time indicators and/or time trends serves as an important control, accounting for costs changing over time that are not captured by the other variables included in the models. This might be as a result of climate change, raw water deterioration, or ongoing work following improvements in service.

To avoid any risk of double counting costs, it would be sensible to discuss whether the real price effect (RPE) component of the efficiency challenge should be removed if a time trend is included in the models directly.

Sample period selection (Chapter 3.4)

Increased time periods for the data panels will help the quality of models. Therefore, we support the use of the full dataset back to 2011-12. However, we suggest Ofwat should not use business plan forecasts in models because it is likely to materially reduce the robustness of models.

Cost adjustment claims (Chapter 4)

Cost adjustment claims are a critical safety valve for achieving an accurate picture of companies' efficient costs. We consider they should be used in a way that reduces risk of model error and allows for costs that cannot realistically be modelled.

Symmetrical adjustments appear to be a useful way of addressing model specification issues. They can help to address known weaknesses that cannot be included in models due to data limitations. However, non-symmetrical claims are also required where costs have not been incurred in the past (here, the model cannot predict the costs, rather than the model distributing the costs incorrectly) or costs relate to only one company (the symmetrical adjustment would be immaterial).

We suggest materiality should be used before adjustments are applied. This will reduce the number of potential adjustments which could otherwise become unworkable.

We would also like to see guidance improved so that it is clear how different types of expenditure can be assessed (particularly the approach to the implicit allowance).

Capital maintenance and asset health (Chapter 5)

We believe that future capital maintenance challenges are a fundamental issue that needs to be explored further.

We do not support using business plan forecasts as inputs into base econometric models, and consider this would materially reduce the robustness of the modelling. We note that the use of a time trend explanatory factor could provide a means of taking greater account of future pressures without the complicating effects of using forecasts as input data. We support more consideration being given to the use of forecast efficiencies when setting the catch-up efficiency challenge, but consider this would need to be carefully managed for it to deliver as intended. Also, we think that Ofwat should not necessarily reject the potential for including asset health variables in models alongside appropriate safeguards.

Cost-service relationship (Chapter 6)

There is a need to acknowledge the relationship between cost assessment and service performance. However, we do not yet believe that the current thinking in this area is sufficiently developed.

For the models to reflect relative efficiency it is important that base costs should relate to maintaining, rather than gradually improving, service.

All levels of service are included (implicitly) in the cost modelling. Therefore, base expenditure should relate to the performance of the efficiency benchmark companies. With an appropriate level of tolerance, this increasingly tends towards the sector average. This supports the view that there are material external drivers of performance that are not captured in cost models.

We think that the identified sector average level of performance should be projected forward on a flat basis (subject to any related enhancement expenditure assumed at PR24) rather than being assumed to follow an ongoing trend. We think that scope for productivity improvements should be assessed separately as part of determining an appropriate frontier shift assumption (as it was at PR19), as this approach would align with the underlying definitions of base and enhancement, and would avoiding the risk of double counting issues arising.

We do not think it is sensible to model performance levels through econometric models at this stage. This would drive too much uncertainty. Ofwat could consider requiring companies to explicitly separate base expenditure required to maintain 2024/25 service standards and expenditure required to meet forecast 2029/30 service levels. This could then be considered together with any policy on continuing efficiency assumptions.

2. Principles of PR24 base cost assessment

Summary of content / discussion set out in consultation document

- Ofwat outline their proposed cost assessment principles for PR24:
  - Consistent with engineering, operational and economic rationale
  - Sensibly simple and transparent
  - Focus on exogenous cost drivers
  - Robust econometric cost models
  - Set a stretching but achievable cost efficiency challenge
  - A coherent cost assessment approach

Consultation Questions asked
1. Do you agree with our principles of base cost assessment?
2. Do you consider any important principles are missing?

Severn Trent Response

We are broadly supportive of the suite of principles identified and strongly agree with principles 1, 2 and 6. We briefly discuss each in turn and make some cautionary comments where appropriate.

Principle 1 - Consistent with engineering, operational and economic rationale

It is critical that models are underpinned with engineering, operational and economic rationale. This is consistent with the framework that we developed, our engagement and our model selection work that we undertook and shared during PR19. We summarise the key points in our response to ***Chapter 2.1*** below as we believe that they remain relevant to econometric model development at PR24. We also use this understanding to base the more detailed comments and suggestions that we make elsewhere in this consultation response.

Principle 2 – Sensibly simple and transparent

We agree that models should be sensibly simple and transparent. Overly complicated models can lead to statistical problems around overfitting, and in some cases reduce interpretability. Some potential ways to address these concerns are outlined in our response to ***Chapter 2.2***, where we propose the use of information criteria which provide evidence for the quality of model fit while marking down models that include too many variables.

Principle 3 – Focus on exogenous cost drivers

We understand the high level principle that models should be seeking to control for the environmental and regional circumstances that are placed upon a company rather than being influenced by the decisions and choices that they make. However, we think that there should be some level of flexibility to consider endogenous variables if statistical tests are satisfied and there is no possibility of perverse incentives being an issue.

We propose that there could be a distinction between variables that are under management control in the short term, which should not be included in the models, and variables that are only under management control in the long term, which could be included.

Principle 4 – Robust econometric cost models

A prerequisite for accurate models is that they are constructed using robust and consistent data. We acknowledge the need for econometric models to be statistically valid and robust. A range of diagnostic tests should be run to provide insight into the performance of different model specifications.

However, we do not believe that statistical tests should then be used as the primary arbiter of selecting one model over another. Instead, the derived information should form part of the evidence for or against a particular model.

We also consider that value can be had from testing the predictive capabilities of models using historical data. Consequently, principle 4 could be extended to 'Robust econometric cost models that are fit for purpose'. That is to say that they both satisfy statistical tests, but also provide adequate forecasts when tested against real data.

In our response to ***Chapter 2.2*** below we set out a more detailed commentary on the use of statistical methods and information, and how we consider that they should be used to inform model development and selection. We then use this understanding when making further comments later in our response.

Principle 5 – Set a stretching but achievable cost efficiency challenge

We have stronger concerns with this principle and make the following comments:

- Fundamental premise and expectation of efficiency challenge – The strength of efficiency challenges should reflect the quality and robustness of the models and wider assessment approaches that have been used. They should not be used retrospectively to arrive to prior expectations of expenditure levels.
- Catch-up efficiency challenges set at previous price reviews and by other regulators in comparator sectors – The 'other sectors' considered must be genuinely comparable, both in nature of activity and in the nature of the price review process for them to be considered with appropriate adjustments. Using the efficiency challenge of other regulators whilst not considering their more general cost assessment processes would violate principle 6 of a coherent cost assessment approach.
- External cost benchmarking analysis – Whilst there could be potential in using external benchmarking, the process would need to be transparent; this would need to include demonstration as to why the other sector could be considered analogous in terms of activity and how incurred costs and efficiencies will be rendered comparable.

We cautiously welcome the consideration of a forward-looking efficiency challenge, believing it could be a pragmatic way to account for future challenges. However, given the forecast costs will necessarily be less robust than the estimated historical costs, the forward-looking efficiency challenge should be less a stringent than one created using outturn costs.

Principle 6 – Coherent cost assessment approach

We very much support the need to develop a coherent and holistic cost assessment approach. Base models, enhancement models, cost adjustment claims, frontier shift and the catch-up efficiency challenge are all important parts of the cost assessment process. Changes to one area should be assessed for their impact on other parts of the process and adjusted accordingly. A key enabler to developing a coherent approach is to specify each of the components with sufficient time to test how they interact.

2.1 Using engineering logic to underpin the construction of econometric models

While models should be coherent from both an engineering and statistical standpoint, we are of the view that the engineering rationale is the most important aspect. It should underpin the selection of all drivers. The statistical validity of these models should then be assessed, and the model rejected if it does not work from a statistical standpoint. Essentially, we should not select a model based purely on its 'goodness of fit' if the variables it uses cannot be supported by sound engineering rationale. Once we have a pool of models for each cost category that satisfy the engineering challenge, we should select between them by assessing their statistical quality.

We refer back to our cost modelling framework published for PR19 which outlined key factors affecting water network+ botex based on expected engineering logic, and our consultation response at PR19 which attempted to do the same for waste.

Our PR19 proposed framework for Water Network+ factors is shown in the following table:

Key underlying factor	Why is it relevant to costs?
1. The distance water has to be transported (which will be affected among other things by the relative locations of customers and sources of water)	Transporting water over longer distances will require more assets that will need to be maintained. Operating costs can also increase with distance (e.g. when more pumping is required).
2. The number of customers to whom water is distributed	More customer connection points will require a greater scale of distribution infrastructure, including water mains, pumping equipment and service reservoirs.
3. The quantity of water that has to/may have to be transported and treated	Peak transportation and treatment requirements will affect capital costs as they will affect the capacity that assets need to provide for. The quantity of water will also affect operating costs (e.g. energy for pumping and chemicals for treatment).
4. The geography (including the topography) over which water has to be transported	Pumping costs will be affected by topography. Geography can affect costs in a range of other ways such as through effects of pipe maintenance requirements (e.g. as a result of corrosion).
5. Opportunities for the achievement of economies of scale in the transportation of water	When customers are more spread out, more assets will typically be required per customer, and opportunities for economies of scale in network assets can be more limited (e.g. more, smaller pumping stations are likely to be required alongside a greater length of lower-diameter water mains).
6. The extent to which transportation activities are affected by congestion issues	Operating in densely populated urban areas can put upward pressure on costs as a result of above and below ground congestion, and the implications this can have on the need for, and complexities and resource requirements associated with, carrying out maintenance work.
7. Opportunities for the achievement of economies of scale in water treatment	There can be significant economies of scale in water treatment, and so the extent to which those economies are achievable in practice can have a material bearing on costs.
8. The extent and forms of treatment that are required given the quality of raw water	Treatment costs will be affected by what forms of treatment have to be provided, and that will depend on the quality of the raw water.
9. Regional differences in relevant input costs	Labour and energy costs can vary between regions in ways that have a material bearing on overall transportation and treatment costs.
10. Service quality variations	Costs can be affected by differences in the levels of service quality that different companies provide. This could include the provision of environmental and related outcomes (such as leakage levels).

11. The significance of other customer characteristics that may affect transportation and treatment requirements	For example, meter penetration levels and the relative size of non-potable demand may affect cost materially.

And our proposed framework for waste network plus that was presented in the PR19 cost assessment consultation is outlined in the following table:

Driver category	Specific example
Scale drivers	The distance sewage has to be transported (network size). The total nutrient load within sewage to be removed before it can be safely returned to the environment.
Network-specific drivers	Challenges of geography and topography within the network. Opportunities for economies of scale in the transportation of sewage (network density). The extent to which transportation activities are affected by 'congestion'.
Treatment-specific drivers	Opportunities for economies of scale in sewage treatment. The extent and forms of treatment that are required (treatment complexity).
Other drivers	Regional differences in relevant input costs. Service quality variations. The significance of other customer characteristics.

In our PR19 cost assessment framework, we further outlined the way in which we could test that these variables had been appropriately considered in the cost assessment framework. This fundamentally asks the question **"Is each factor captured in the model specification?"**:

Potential response	Details/comments
a. Yes, directly	The factor is explicitly captured through the inclusion of a given explanatory variable.
b. Yes, indirectly	The factor is picked up as other explanatory variables provide a sufficiently close proxy, either individually or in combination.
c. No, but factor not sufficiently material	The factor has been identified as not having a sufficiently material effect on costs for inclusion to be necessary.
d. No, but ex-model adjustment to be made	A separate, targeted adjustment has been, or will be, made outside of the model (either a pre-modelling or post-modelling adjustment) so that the factor does not need to be captured through the model specification.
e. No, factor not captured	The factor looks likely to have a material bearing on costs but is not picked up in either the model specification or in an adjustment outside the model.

For the potential responses 'a' to 'd', the factor is either adequately captured as part of the cost assessment process or immaterial, but if the potential response is 'e', an amendment should be made to some part of the process to capture the associated costs.

Throughout our response to ***Chapter 3*** we suggest several changes to variables that were considered in the PR19 models, and we suggest some variables that were not included in the PR19 modelling that we believe would be worth considering at PR24. We have outlined the rationale behind these variables and have commissioned consultancy support to look at the engineering logic behind our suggestions.

2.2 Using statistical methods and information to improve/validate models that have been identified based on engineering logic

In line with principle 4, the use of statistical and econometric techniques is very important in producing robust econometric models. We are broadly supportive of Ofwat's approach as outlined in section 3.6 of its consultation document.

This chapter relates to section 3.5 and section 3.6 of the consultation document. Consequently, the following consultation questions are also relevant here:

Consultation Questions asked
Model estimation method
12. Do you agree that we should maintain the use of random effects to estimate our wholesale base cost models at PR24?
Model selection process
13. Do you agree with our proposed model selection process?

We support the retention of Random Effects as the estimator of choice. It remains fit for purpose and therefore it is sensible to retain a level of consistency with the PR19 analysis where there is no strong reason to change.

After deriving a range of appropriate models on engineering grounds, overarching tests of model regression fit and variable confidence should be reviewed to challenge the underlying specification of models. Subsequently a series of more specialised statistical tests and measures can be also used to refine and ensure the most appropriate models are selected for use.

We provide a number of additional tests and metrics that were not used at PR19, but that we consider to be valuable in aiding model selection. These focus on the model 'goodness of fit' while avoiding problems with overfitting, and the quality of the model in terms of its predictive power, which is the fundamental purpose of producing these models.

Standard statistical measures – model regression fit and explanatory variable confidence

The standard test of **statistical performance** in a regression model is the R^2 value. High R^2 values are important for model robustness. However, they are only of limited benefit for testing the appropriate wider specification of model coefficients. This is because most of the explanatory power of a model will be controlled by the scale driver. This means that the impact of other key variables such as density and complexity will be relatively insensitive. A potential remedy may be to normalise models by the primary scale driver to allow for examination of the fit of the other variables. A modified R^2 value may also be a more useful metric as it accounts for the number of observations and explanatory variables.

The **significance of explanatory variables** measured by p-values calculates the probability of the relationship being observed under the null hypothesis (i.e. as a result of noise or chance). It is appropriate to expect models to have an appropriate level of confidence (ideally <0.05, or at most <0.1). However, we think that where a variable supported by strong engineering logic appears insignificant, this should lead to further review of that underpinning evidence or testing of the interaction of variables (e.g. collinearity) rather than removal per se.

Considering more specialist statistical measures

More specialist statistical measures and tests can be used to consider potential modelling issues such as overfitting, multicollinearity and misspecification. The summary implications of these are as follows:

- It is important to consider the fit of models whilst also limiting the potential problems of **overfitting**. Overfitting occurs where variables are added to a model because they help to describe the distribution of data in the modelling data set. However, where the variables are not causal drivers of cost, they will reduce rather than improve the ability of models to forecast future costs accurately. We believe that the Akaike and Bayes information criteria (***Appendix 8.8***) are the best diagnostic tests to use for this.
- Simple econometric cost models assume a series of independent linear relationships between cost (the dependent variable) and identified cost drivers (the explanatory variables). If the reality is more complex (i.e. there are non-linear relationships or interactions between variables), the functional form of the model will likely be mis-specified leading to a biased model. This is considered in **misspecification tests**. We believe that the RESET test (***Appendix A.8***) is the best diagnostic test to use for this.
- A range of empirically supported cost drivers may be available for potential inclusion into cost models. However, issues can arise if there is strong correlation between the independent cost drivers. This is called **multicollinearity**. It can often lead to spurious and unexpected coefficients where models cannot accurately allocate causal relationships between the drivers. Whist this may not necessarily be a problem for accurately determining total costs, it makes models difficult to interpret and challenge. We believe that the Variance Inflation Factor (***Appendix A.8***) is the best diagnostic test to use for this.

Several of these statistical tests will require use of a different estimator relative to the preferred standard Random Effects estimator. These are alternative methods for estimating the same model coefficients. For example, the Maximum likelihood random effects estimator (MLE) is required for calculating the likelihood function that is the basis of the Akaike and Bayes information criterion (AIC and BIC), and OLS is required for calculating the variance inflation factor (VIF).

The use of different estimators will give slightly different estimates of the coefficients, but in practice for model development purposes these differences will be minor. Therefore, we strongly recommend that these alternative estimation methods should be run to enable the analysis. Once models have been appropriately tested and selected, a random effects estimator can then be used to calculate expenditure requirements in line with Ofwat's preference.

Further information on AIC, BIC, RESET and VIF is shown in Appendix ***A.8***.

Testing the predictive capability of models – Train-test split evaluation

The main purpose of creating econometric cost models is to predict future costs. Therefore, we think that it would be very sensible to test proposed models on actual values that are independent to those used in the construction of the model coefficients. This is called a 'train – test split evaluation'.

Models should be developed and refined using all the available data. This will identify a preferred set of explanatory variables to be tested. To properly assess the predictive power of identified models and ensure they are giving a fair cost allocation based on future variables, the available data panel would then need to be split into two sets:

- the training set used to fit the models (the start of the data panel), and
- the test set that will be used to assess them (the final couple of years of the data panel).

After the model coefficients have been recalculated following fitting to the training set, we then estimate the dependent variable for the data that has been withheld as the test set. The predicted values are then compared to the actual values to get an understanding of how well the model works in a predictive capacity. It's important to note that we should not seek to improve the performance of these refitted models. This would simply risk overfitting on the test set which will suffer from the same problems as overfitting on the full data panel.

For our purposes the best candidate to test model fit is probably Mean Absolute Percentage Error (MAPE). There have been some criticisms of MAPE due to it being potentially biased towards underestimation. But given the values that we're forecasting, there shouldn't be any issues associated with this.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|$$

Where we have presented model specifications in this document, we have identified the MAPE for using a test set of the final two years of the panel.

To extend the analysis further, a cross validation should be undertaken. This would typically consider several randomly created training sets and test sets. However, since our panel data is constrained by the time component, this is not feasible. Instead, we could use a more programmatic approach. For example, where our first training set runs from years 1-5, testing on years 6 & 7. Then our second training set runs from years 2-6, testing on years 7 & 8, and so on until we run out of data. We would then average out the MAPE across all models and select the models that give the lowest values.

This approach would guard against overfitting to a particular test set and mean that our model works consistently over time. However, where time indicators are included in model specifications (as discussed in ***Chapter 3.3***) this may create issues so would need to be used with caution.

3. Approach to wholesale base cost modelling at PR24

3.1 Scope of wholesale modelled base costs

Summary of content / discussion set out in consultation document

- Ofwat's starting position is that the scope of modelled base costs will be the same as at PR19, i.e. Botex+, base costs with some elements of growth.
- Ofwat is open to the removal of growth costs from the models, but are of the view that there will need to be substantial improvements in data to enable the specification of robust standalone growth models.
- Ofwat has introduced some new developer services data lines in the 2020-21 APR, which it plans to continue collecting in APRs moving forward.

Consultation Questions asked

Scope of wholesale modelled base costs

3. Do you consider the scope of wholesale modelled base costs should be amended at PR24? If so, please explain how the potential amendment/s to wholesale modelled base costs can be justified based on our proposed assessment framework.
4. Would you recommend collecting additional data in relation to growth expenditure (cost and/or cost driver data) to improve cost assessment at PR24? If so, what additional data would you recommend collecting? Please provide definitions alongside suggested data additions.

Severn Trent Response

- If separate robust models can be formulated then growth costs should be removed from the base modelled costs, i.e., we should move from a Botex+ back to a Botex model scope.
- Network reinforcement costs are particularly lumpy. They are driven by specific local conditions which can't be modelled by single company-wide variables, and are therefore particularly problematic.
- If growth costs cannot be removed from the models, it would be pragmatic to consider adding explanatory variables aimed at describing those costs to model specifications.
- We do not agree with the addition of further costs into Botex+ models. Any additional costs are likely to be lumpy and will further add to the inherent 'noise' in models making it harder for Ofwat to identify efficient costs.

Supporting analysis undertaken on the treatment of growth costs in models

The assessment of developer services / growth costs are complex. The amount of attention given, and the number of iterations that were made in attempting to get to an appropriate solution at PR19 bear this out.

From a theoretical or engineering perspective, we do not consider that it is sensible to include developer services/growth costs in econometric models which are specified to understanding drivers of base cost. The costs do not closely relate to ongoing base expenditure and there is clear evidence to suggest that costs vary regionally in line with different growth rates and specific regional circumstances.

At PR19 we strongly argued that company growth cost and revenue data was not sufficiently consistent to have confidence in separate developer services models. Consequently, we supported the pragmatic decision to expand the Botex models to also account for growth costs (Botex+).

Since PR19, through a series of data requests and the current RAG4.10 consultation (November 2021), Ofwat has worked hard to improve the way in which developer services / network growth costs are reported. This may support the development of more robust separate growth models enabling the dependent variable in econometric models to return to reflecting just base expenditure (Botex). However, it is currently too early to use the new data in model development to offer an informed opinion at this stage.

The analysis below in 'Testing the treated water distribution model specification for varying inclusion of growth' considers the impact of changing the extent to which growth costs are included in the dependent variable of Botex+ econometric models. Whilst we consider that this could help with the specification and robustness of assessing base costs in econometric models, we acknowledge that an appropriate assessment of costs removed from econometric models will also be required (i.e., through separate models or deep dives of evidence submitted in business plans). Therefore, if this cannot be delivered using the updated developer services cost information that Ofwat will have, a Botex+ approach as per PR19 may remain a pragmatic solution.

Our analysis here focuses on the impact of growth expenditure on water models. A similar logic could be expected in waste models. This would relate to waste growth expenditure and sewer flooding costs which were included in the scope of the PR19 Botex+ models. In all cases, where costs with distinct cost drivers are included in models without associated explanatory factors, this is likely to increase noise and therefore reduce the predictive power of models.

The engineering logic for drivers of growth costs and their appropriateness for modelling

In the table below, we describe the specific attributes of the different types of developer services (growth) costs. Costs most suitable for modelling are typically high volume activities to minimise the difficulties of lumpiness and driven by factors that are likely to be applicable to all companies which can be tracked by clear and consistent metrics.

Cost (location)	Description and Location of costs	How likely is this to be influenced by local circumstances?	Subject to contestable market?
New connections (on-site)	Connecting properties to the existing network or new requisition main including coms pipes, boundary boxes and meters.	Activity is largely repeatable and high-volume. Will be impacted to an extent by the type of properties being connected.	Mostly[1]
Requisitions (on-site)	Laying mains on construction sites that will eventually connect new connections to the existing network.	Activity is repeatable. Will be impacted by the specific attributes of construction sites but should broadly tend to an average if volumes are sufficiently large.	Yes
Network reinforcement (off-site)	Modifications to the existing network to ensure that growth can be accommodated and levels of service maintained. Could include upsizing mains, construction of DSRs or pumping stations.	Highly influenced by local circumstance. Costs are highly variable depending on existing asset context. Only likely to reflect growth rates over very long timescales and could be complicated by the growing influence of NAVs.	No

[1] The connection to exiting network – 'source of water connection' – is undertaken by incumbents.

At a high level, fundamental drivers of developer services / growth costs can be considered as:

- The volume of new connections (e.g., dependent on regional demand for new housing)
- The attributes of development sites (e.g., length of requisitions /density of housing, economies of scale on large sites, ground conditions faced, types of new connections)
- The existing network capacity at the specific location (e.g., need for / extent of network reinforcement)
- The amount of self-lay and NAV activity (much greater impact on gross rather than net costs)

Network reinforcement activity is highly localised and sporadic. It is not likely to be sensitive to typical base expenditure focused cost drivers – Identical development sites can have very different reinforcement costs depending on the location and timing of the development. For example, in some cases, new development can lead to no network reinforcement where there is sufficient local network headroom at the time. This is already acknowledged in the calculation of the infrastructure charge where total reinforcement costs are averaged over a long time period. Consequently, from an engineering logic perspective, these are not likely to be reflected by typical base expenditure explanatory factors.

The on-site costs will legitimately exhibit cost variance between development sites (as described above). However, given that the assets and activities are similar and of relatively high volume, we would anticipate that the costs are more likely to behave like base expenditure than network reinforcement.

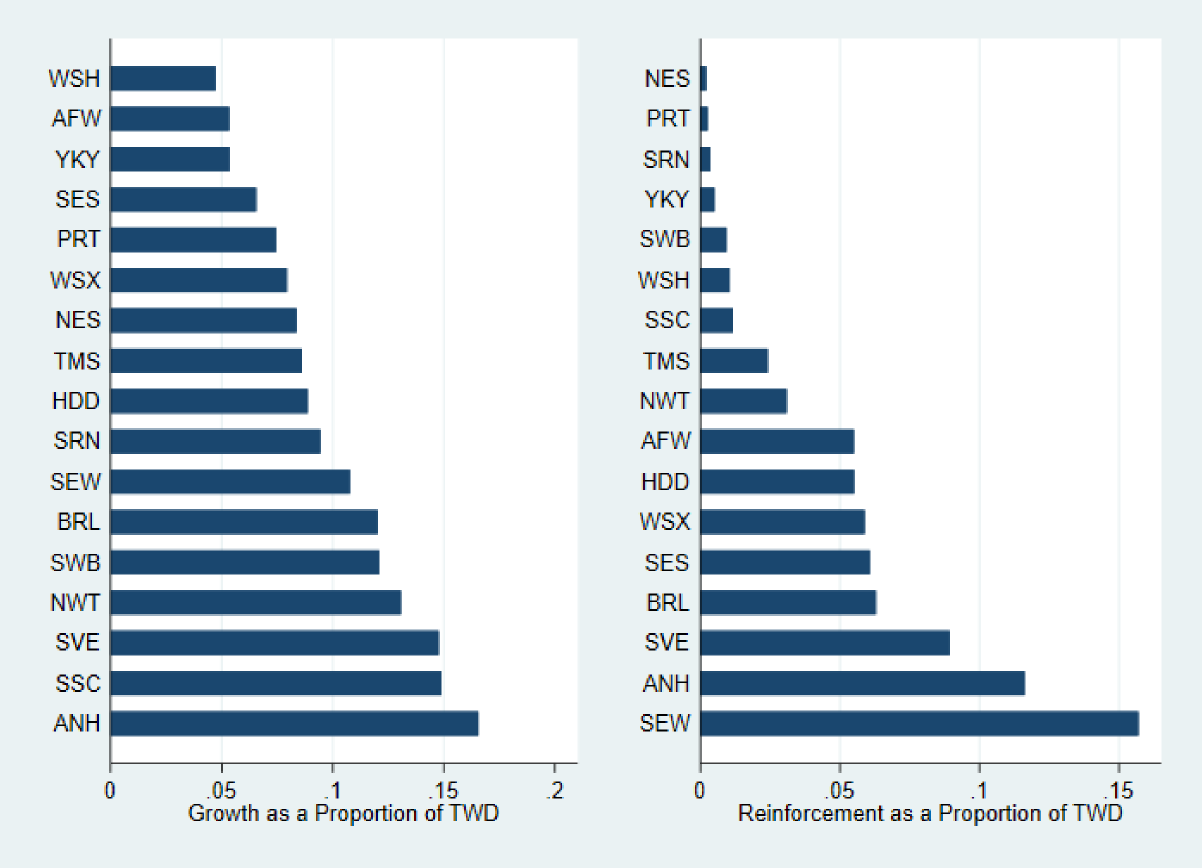
Given that there are no developer services / growth explanatory factors in the current econometric model, growth costs would need to be sensitive to the existing scale, density and complexity/ explanatory factors. There is limited engineering, operational or economic rationale to link growth costs to density or complexity drivers. Therefore, where growth costs might be more suitable for modelling (i.e., on-site rather than off site costs), the explanatory power will currently need to come from the model scale driver. However, this would require growth costs to be of a similar proportion to base costs between companies (i.e., growth activity is relatively equal across the country). The graphs in ***Figure 3.1.1*** show that this is not the case.

The interaction of the two graphs also provides support to our assertion that network reinforcement costs do not strongly follow the wider growth costs (due to the local and contextual circumstances affecting reinforcement activity).

We also note that this information highlights potential data consistency issues (in some cases reinforcement is greater than growth even though it should be a subset). We hope that the more granular reporting in future APRs will help to alleviate this.

For the above reasons, we consider that, from an engineering logic perspective, it would be sensible to include a specific growth explanatory factor in models where potentially modellable growth costs (most justifiably on-site costs) are to be included in the scope of econometric botex+ models.

Figure 3.1.1 – Total Growth and Reinforcement costs as a proportion of Treated Water Distribution costs. HDD and SVE were combined with DVW and SVT respectively for the purposes of creating these charts.



Testing the treated water distribution model specification for varying inclusion of growth

Given the above discussion on engineering logic for growth costs, we have tested the following modifications to the existing TWD model:

- A model that accounts for all base and growth costs (on-site and off-site). This is the current botex+ model scope. No separate growth models or deep dives are required. We can either:
  - Retain existing explanatory factors (baseline). All growth costs to be explained by existing base expenditure explanatory factors (scale, density, network complexity). Model specification 1.
  - Add a growth driver to specification (new properties). Additional explanatory factor to improve explanation of growth costs. Model specification 4.
- A model that removes Network Reinforcement costs (i.e. the most unpredictable growth costs). This is a reduced botex+ model scope. A separate assessment of network reinforcement would be required. We can either:
  - Retain existing explanatory factors. Remaining on-site growth costs to be explained by existing base expenditure explanatory factors (scale, density, network complexity). Model specification 2.
  - Add a growth driver to specification (new properties). Additional explanatory factor to improve explanation of remaining on-site growth costs. Model specification 5

- A model that removes all growth costs (New connections, requisitions, Network reinforcement). This is a botex model scope. A separate assessment of growth costs would be required (separate model or deep dive assessment). We have retained existing explanatory factors. Explanatory factors do not need to explain any growth costs. Model specification 3.

Model results and interpretations

Figure 3.1.2 – Regression coefficients (p-values) from various TWD/Growth specifications, using a panel of data from 2011-12 to 2019-20.

Variable	TWD (as was – onsite + offsite growth included)	TWD No Reinfor-cement (includes on-site)	TWD No Growth (botex only)	TWD (as was – onsite + offsite growth included)	TWD No Reinfor-cement (includes on-site)	Reinforc-ement only	Growth only
Model specification	1	2	3	4	5	6	7
lnwedensitywater	-3.1969 (0.000)	-3.0257 (0.000)	-3.2250 (0.000)	-2.7563 (0.000)	-2.6518 (0.000)	-2.8950 (0.318)	-2.1533 (0.475)
lnwedensitywater2	0.2562 (0.000)	0.2455 (0.000)	0.2580 (0.000)	0.2239 (0.000)	0.2182 (0.000)	0.1918 (0.341)	0.1728 (0.386)
lnlengthsofmain	1.0620 (0.000)	1.0506 (0.000)	1.0486 (0.000)	1.0402 (0.000)	1.0310 (0.000)	0.6385 (0.000)	0.8787 (0.000)
lnboostersperlength	0.5504 (0.000)	0.5886 (0.000)	0.5740 (0.000)	0.5819 (0.000)	0.6216 (0.000)	-0.5979 (0.479)	-0.2612 (0.772)
prop_newconnections	-	-	-	0.3587 (0.000)	0.3503 (0.000)	-	-
constant	6.1092 (0.000)	5.6681 (0.000)	6.3394 (0.000)	4.6243 (0.009)	4.4863 (0.000)	3.3153 (0.704)	-1.2150 (0.901)
R^2	0.9623	0.9622	0.9632	0.9651	0.9635	0.2848	0.5586
RESET	0.0481	0.0613	0.1013	0.1639	0.1513	0.0048	0.0544
MAPE 2018-20	20.45%	21.18%	19.82%	20.12%	20.18%	93.35%	55.50%

Figure 3.1.3 – Efficiency values based on the regressions without 'prop_newconnections' in Figure 3.1.2. Efficiency changes are relative to the original specification.

Company	TWD (model1)	Rank	TWD No Reinforcement (model 2)	Rank	Efficiency Change	TWD No Growth (model 3)	Rank	Efficiency Change
SWB	0.80	1	0.82	2	2.50%	0.77	1	-3.75%
PRT	0.80	2	0.82	1	2.50%	0.81	3	1.25%
SRN	0.81	3	0.82	3	1.23%	0.79	2	-2.47%
NWT	0.89	4	0.91	4	2.25%	0.86	4	-3.37%
WSX	0.92	5	0.93	5	1.09%	0.91	6	-1.09%
SSC	0.95	6	0.97	6	2.11%	0.88	5	-7.37%
TMS	0.98	7	1.03	9	5.10%	1.01	8	3.06%
SVE	1.02	8	0.99	7	-2.94%	0.98	7	-3.92%
YKY	1.03	9	1.06	10	2.91%	1.07	10	3.88%
HDD	1.05	10	1.06	11	0.95%	1.03	9	-1.90%
SES	1.05	11	1.01	8	-3.81%	1.08	11	2.86%
WSH	1.09	12	1.13	13	3.67%	1.14	13	4.59%
NES	1.10	13	1.15	14	4.55%	1.13	12	2.73%
AFW	1.10	14	1.07	12	-2.73%	1.15	14	4.55%
BRL	1.27	15	1.23	16	-3.15%	1.22	17	-3.94%
SEW	1.29	16	1.18	15	-8.53%	1.22	15	-5.43%
ANH	1.33	17	1.28	17	-3.76%	1.22	16	-8.27%

Figure 3.1.4 – Efficiency values based on the regressions with 'prop_newconnections' in Figure 3.1.2. Efficiency changes are relative to the original specification.

Company	TWD (model 4)	Rank	Efficiency Change	TWD No Reinfor-cement (model 5)	Rank	Efficiency Change
SWB	0.73	1	-8.75%	0.76	1	-5.00%
PRT	0.79	2	-1.25%	0.80	2	0.00%
SRN	0.79	3	-2.47%	0.80	3	-1.23%
NWT	0.90	5	1.12%	0.92	5	3.37%
WSX	0.85	4	-7.61%	0.86	4	-6.52%
SSC	0.94	6	-1.05%	0.95	6	0.00%
TMS	1.01	8	3.06%	1.04	10	6.12%
SVE	1.06	11	3.92%	1.04	9	1.96%
YKY	1.04	10	0.97%	1.07	11	3.88%
HDD	1.10	12	4.76%	1.12	13	6.67%
SES	1.03	9	-1.90%	0.99	8	-5.71%
WSH	1.14	14	4.59%	1.19	16	9.17%
NES	1.10	13	0.00%	1.15	15	4.55%
AFW	0.98	7	-10.91%	0.96	7	-12.73%
BRL	1.19	15	-6.30%	1.15	14	-9.44%
SEW	1.20	16	-6.98%	1.10	12	-14.73%
ANH	1.24	17	-6.77%	1.20	17	-9.77%

The main inferences we draw from our modelling of growth and reinforcement costs are as follows.

Company efficiency performance changes systematically based on the relative proportion of growth costs it has (***Figure 3.1.3***). This is true for both the network reinforcement growth and total growth. Where companies have a higher proportion of growth, they become more efficient following the removal of growth costs from the model scope and vice versa. Were growth costs explainable by the current explanatory variables, we would expect that efficiency levels would remain roughly constant. The systematic changes in efficiencies instead suggest growth costs are simply adding noise, with high (low) growth proportion businesses having their costs systematically underestimated (overestimated).

The addition of 'prop_newconnections' (the proportion of year end connections that were newly connected in the given year) as an explanatory factor ensures that efficiencies do not change in such a systematic way (***Figure 3.1.4***). However, adding such an explanatory factor is not likely to be sufficient to give a full picture of the costs associated with even onsite work. This this because the additional variable is likely to account for the scale of growth rates, rather than reflecting the difficulty/complexity of the activity being delivered.

We can see that when reinforcement or growth are used as the dependent variables (***Figure 3.1.2,*** *model specifications 6 and 7*), the models reduce significantly in quality. Only the scale variable remains significant and the R^2 values become very low. This further highlights that there is very little variation in growth costs that is explained by the existing explanatory variables in the model.

We have not attempted to add an explanatory variable to reflect network reinforcement. It is particularly difficult to model given its intrinsic relationship with local network capacity at the time of the new connections. We do not think that there is an obvious way to construct a company-wide variable to account for these factors.

Therefore, based on our analysis we would encourage Ofwat to explore the following options to better account for growth costs:

Scope of base econometric model	Model specification	Off model analysis
Base (Botex)	No change with respect to growth	On-site costs: explore if separate model can be developed to an appropriate level of robustness Off-site costs: Deep dive of network reinforcement business case
Base + Onsite growth (Reduced Botex+)	Add an explanatory factor sensitive to growth (further analysis needed to identify sensible drivers)	Off-site costs: Deep dive of network reinforcement business case

3.2 Wholesale base cost modelling suite

Summary of content / discussion set out in consultation document

- Ofwat is intending to remove Bioresources from the Wastewater suite and determine its revenue as a standalone model.
- It is then intending to add a Network Plus Wastewater (NPWW) model to the suite as a replacement for the old Bioresources Plus (BRP) model to consider wastewater costs at a higher level of aggregation.
- Ofwat is not considering breaking down Water Resource Plus (WRP) into its component parts (Water Resources, Raw Water Distribution, Water Treatment) unless robust models can be proposed for these elements of cost.

Consultation Questions asked

Target modelling suite

7. Do you agree with our proposed target wholesale base cost modelling suite at PR24?
8. Do you consider it would be worthwhile attempting to develop wholesale wastewater network plus models for PR24? If so, do you propose any potential wastewater network plus cost model specifications to consider?

Severn Trent Response

- We are broadly supportive of the proposed modelling suite.
- The addition of a Network Plus Wastewater model is sensible given the removal of Bioresources from the model scope.
- We recognise there is difficulty in the creation of Water Resources and Raw Water Distribution models, and although we would welcome robust models for these cost elements, we believe that retaining Water Resources Plus models at PR24 is pragmatic.

We support the development of models at different levels of aggregation. This should help to expose and account for potential interactions and trade-offs across the value chain and allow more specialised but powerful explanatory factors to be used. In line with our earlier comments on engineering logic and statistical performance, the level of aggregation should also reflect the likelihood of developing coherent and robust models. Therefore, it appears pragmatic not to pursue separate water resources models further. However, there feels like there is a greater potential to sensibly model water treatment costs separately. Alongside water resources plus models this might provide additional support to understanding the boundary between the water resources and water network plus price controls.

We support the intention to develop a Wastewater network plus model for use at PR24. This scope was missing at PR19 but we consider that this is a logical higher level of aggregation particularly given the potential trade-offs which may exist between waste collection and treatment. We have started to consider potential model specifications for wastewater network plus models. This is set out in ***Appendix A.1***.

An effective wastewater network plus model will also be important given the proposed changes to the assessment of Bioresources costs. However, in line with principle 5, Ofwat should be mindful of the need to ensure an appropriate level of robustness across all costs. We suggest there is also a need to acknowledge potential trade-offs between bioresources and network plus wastewater when setting efficiency challenges. Deriving a robust bioresources model could be particularly challenging given the difficulty of modelling bioresources activities coupled with the additional aspirations that Ofwat have in this area (i.e., the inclusion of more enhancement expenditure; depreciation rather than capex; and financing costs). We will discuss this in more detail in our response to the separate Bioresources consultation.

3.3 Wholesale base cost drivers and explanatory variables

Summary of content / messaging in Ofwat's consultation document

- Ofwat identify four key categories of cost to capture wholesale costs – scale, density, topography and complexity.
- Four variables of concern are discussed:

  - Average pumping head as a topography explanatory variable
  - Economies of scale in wastewater treatment, as proposed by Anglian
  - Regional wages and capital stock, which Ofwat do not wish to consider further
  - The use of time trends
- Ofwat welcome feedback on the drivers listed above, as well as any other potential areas of concern.

Consultation Questions asked

Cost drivers and explanatory variables

1. Do you think we should reconsider the inclusion of APH in the wholesale water base cost models at PR24? If so, should it be a substitute for, or additional to, booster pumping stations per length of mains?
2. Should we consider replacing the existing 'load treated in size band 6' variable with 'load treated in band 8 and above' in the relevant wholesale wastewater base cost models?
3. Please provide detailed proposals for any additional / alternative cost drivers and explanatory variables we should consider at PR24, including clearly defined data requirements that would need to be collected from companies.

Severn Trent Response

We have considered a series of areas where we consider that there is potential to improve models. These are:

- **Potential use of composite scale drivers** – Consideration should be given to using Principal Component Analysis to allow a richer view of how scale drives cost. We note that Ofgem developed and used a composite scale variable for its RIIO-GD2 econometric benchmark modelling.
- **Better accounting for the effects of operating in areas of differing population density –** Engineering logic does not support the quadratic weighted average density term. More granular analysis of density data should be used to allow for a more representative view of density (either in models or through a symmetrical cost adjustment).
- **Improvements to water treatment complexity drivers –** We are concerned that the treatment complexity bandings provide an inadequate view of costs. Therefore, further consideration should be given to the way in which treatment costs interact with the definition or aggregation of complexity descriptors.
- **Accounting for both operating and capital cost drivers in water network complexity –** Engineering logic supports the use of both Average Pumping Head (APH) and Boosters per Length of main (BPL) in the model. We do not believe that the discussion in this area should focus on APH replacing BPL, but rather whether APH should be considered as an additional driver subject to data quality.
- **Better reflecting water resource cost drivers in model specifications –** We consider that the Water Resources component of Water Resources Plus is contributing to model noise in the absence of Water Resources variables. Consequently, appropriate Water Resources drivers need to be identified to obtain a robust Water Resources Plus specification.
- **Considering potential changes to wastewater treatment economies of scale thresholds –** Any amendment to the thresholds should be supported by strong engineering logic rather than an argument based purely on sector wide marginal costs. There is a need to ask *how* larger works

reduce costs and at what size we start to see diminishing returns given the economies of scale they facilitate.

- **Accounting for legitimate changes in cost over time –** There is a coherent case for including time variables, whether in the form of indicators or a trend to account for climate change, raw water deterioration or ongoing work following improvements in service. It would be sensible to then consider whether the RPE adjustment to the frontier efficiency challenge would need to be removed to protect customers from paying twice for any price increases which influence the time coefficients. In the long run, without the inclusion of these time controls, costs may be overestimated or underestimated and other coefficients weakened.

These are each considered in turn below.

In **Chapter 2** we emphasised our view that engineering logic should be the primary factor considered when developing potential model specifications. Therefore, we have commissioned consultancy work to review and challenge the engineering logic that we are assuming behind each of the changes we are proposing. We hope that this will improve the level of confidence in potential model changes and will seek to disseminate the information as appropriate.

Scale drivers

It is not always clear which scale driver should be used in model, and there may be arguments for the inclusion of several scale variables. For example, in the wholesale water models, properties were used as the scale driver – more properties means more demand. However, there is a logic for including length of mains given the maintenance costs involved and its inclusion in the TWD regression which makes up the bulk of wholesale costs. Similarly a case can also be made for using population – the higher the population the higher the demand exposing different demographic makeups.

One issue often placed against composite scale drivers is how to determine the weightings for each of the variables identified. Principal Component Analysis (PCA) provides a solution to resolve this issue. It calculates the weightings of normalised variables (principal components) that capture most of the joint signal from the individual variables. The process is discussed in more detail in ***Appendix A.7***. The first principal component – which will generally capture upwards of 98% of the signal for the collective scale variables – can then be included as the scale variable in models.

It is not necessary to include the same principal component in all model specifications. For example, there is less logic supporting inclusion of lengths of main in the Water Resources Plus (WRP) specifications. But it still might be prudent to form a principal component from properties and population since both can be argued to have a sizeable impact on water usage. We can create separate principal components for each specification based on the engineering logic for the particular model scope.

PCA will also reveal secondary and subsequent principal components which will also help to describe the distribution of the data. Where there is clear justification and logic supporting these subsequent terms, they can also be used as explanatory factors in models. This is not likely to be the case where only scale drivers are used to form the principal components. Given they are highly correlated most of the signal will be captured by the first principal component.

A TWD regression that uses the first principal component from *lnlengthsofmain*, *lnproperties* and *lnpopulationwater* is shown in ***Figure 3.3.1*** alongside the original specification. Against each of the measures of model quality, the composite scale variable performs better than just *lnlengthsofmain*. Essentially it avoids the problem of companies being undervalued on scale.

Engineering logic can be put forward linking costs to each of the scale drivers. Whilst lengths of main is an obvious cost driver, connections to the network from properties generate weak points where failures are likely to occur, while population proxying demand can increase deterioration of the network due to higher pressures running through the network.

Figure 3.3.1 – Regression coefficients (p-values) for the original TWD specification, and a specification with a scale principal component in place of the original lengths of main scale variable. These use a panel of data from 2011-12 to 2019-20.

Variable	Original TWD Model	TWD with Scale Principal Component
lnwedensitywater	-3.1969 (0.000)	-2.7153 (0.000)
lnwedensitywater2	0.2562 (0.000)	0.2071 (0.000)
lnlengthsofmain	1.0620 (0.000)	-
pca_scale	-	0.5968 (0.000)
lnboostersperlength	0.5504 (0.000)	0.5808 (0.000)
constant	6.1091 (0.000)	15.4672 (0.000)
R^2	0.9623	0.9657
AIC	-100.87	-105.60
BIC	-79.43	-84.16
RESET	0.0481	0.0692
MAPE 2018-20	20.45%	19.27%

It might appear that composite scale variables are harder to interpret and consequently less desirable under principle 2 (simplicity and transparency). However, the coefficient on the principal component can be decomposed into coefficients for each of the variables included in the analysis. Therefore, the composite variable can be considered as including all the scale drivers in the regression without the problems associated with multicollinearity. This decomposition process is also outlined in ***Appendix A.6***.

Conclusion

We recommend that principal component analysis is considered to avoid the issue of companies incurring costs from material scale drivers that are not included in existing model explanatory factors.

Density

We do not believe that the current weighted average density measure meets its intended purpose, nor is the use of local authority population data to calculate the measure providing consistent information across all companies.

The underlying premise of including weighted average density and its quadratic term (weighted average density^2) was to reflect the engineering logic that costs are elevated at both extremes of density and sparsity. It is true that the two terms allow for a U-shaped distribution of weighted average density with respect to cost. However, the data used does not appear to satisfy these engineering expectations:

- The calculation of weighted average densities is problematic:
  - The size/ level of aggregation of the local authorities in company supply areas will mathematically influence the calculated weighted average values.
  - Weighted averages skew values towards the most populous areas served (generally also the densest areas) irrespective of whether they are typical of the wider supply area. Consequently,

it is not likely to be a good representation of the sparsity of a company's supply area away from the urban population centres which are dominating the calculated value.

- Density is likely to drive costs at a granular sub-company level. Providing one value of density per company describes a single point towards the centre of the distribution. However, the cost pressures are at their greatest where there are areas of extreme density and/or sparsity. When averaged out at a company level, both these cost pressures will be significantly diluted and poorly reflected in an averaged figure.

We have commissioned consultancy support to review and evidence the engineering logic which states that both high density and sparsity areas increase cost pressures. We are also seeking evidence and justification for suitable density thresholds for 'high density' and 'high sparsity' measures.

Issues with weighted average density data – Size of local authorities

The construction and use of weighted average density data appears problematic; mathematically the number/size of local authorities has a material bearing on the density calculated. The more disaggregated the local authorities within a company supply area, the higher the weighted average density will be. This is illustrated in ***Figure 3.3.2***.

Figure 3.3.2 – Weighted average density increases with disaggregation of areas.

Unit	Population	Area	Density (*Weighted Total*)
Single area			
A	100	1	100
Total	*100*	*1*	*100*
Some disaggregation – e.g. large combined authorities			
A	60	0.5	120
B	40	0.5	80
Total	*100*	*1*	*104*
Increased disaggregration – e.g. smaller local authorities			
A	60	0.5	120
B	30	0.1	300
C	10	0.4	25
Total	*100*	*1*	*164.5*
Most disaggregation – e.g. postcode sectors			
A	50	0.1	500
B	30	0.1	300
C	10	0.4	25
D	10	0.4	25
Total	*100*	*1*	*345*

Issues with weighted average density data – skew to the density of population centres

The weighted average density metric is skewed heavily towards the most populous, and in general the densest, areas that the company serves. ***Figure 3.3.3*** shows that the 2011 weighted average density for multiple companies is significantly above the 75th percentile of areas served. For some companies the weighted average is even considered an outlier.

Some of this distribution will also be a function of the differential sizes of local authorities (as discussed above). Therefore, we have attempted to better show this effect by completing a similar density analysis at a postcode sector level (***Figure 3.3.4)***. This shows the true distribution of densities across the supply area given that the observations are geographically much smaller and of a more equal population size. It reveals that the weighted average density produced lies consistently within the 75th percentile of postcode sector areas for all companies irrespective of the distribution or extent of sparse areas that they serve.

This means that the density terms in the PR19 models are not actually accounting for the increased costs of operating in rural areas. Rather it suggests that there are higher costs associated with having the least dense urban areas, which is not in line with the engineering expectation. It also targets a level of density that is not actually representative of most of the area that a company needs to supply.

The mathematical concern alongside the skew towards higher population, denser areas suggest the metric using local authorities is fundamentally not comparable between companies given the different levels of disaggregation between them.

Figure 3.3.3 – Boxplot for companies' Local Authority population densities. The large black point for each company is the weighted average density.

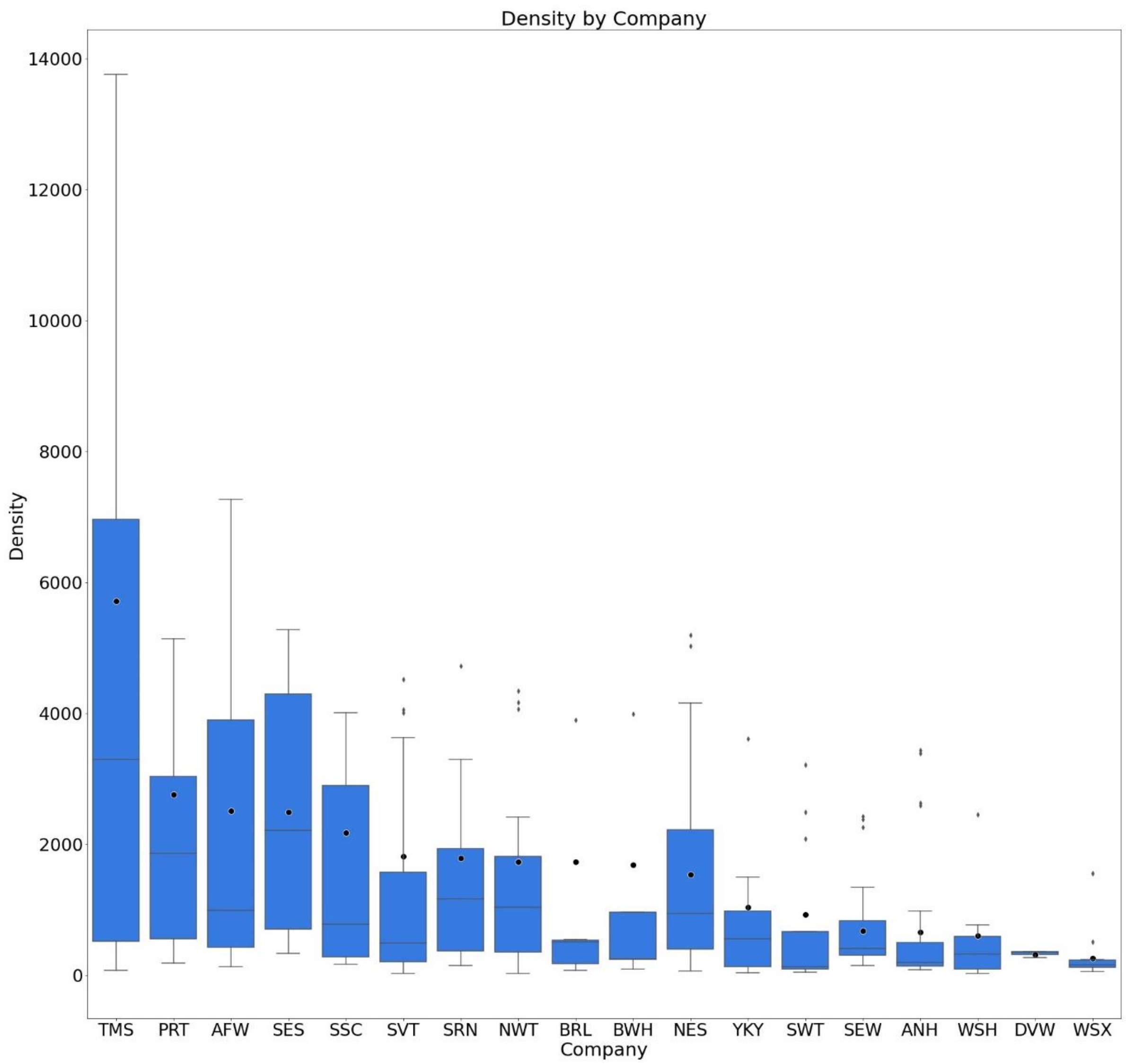
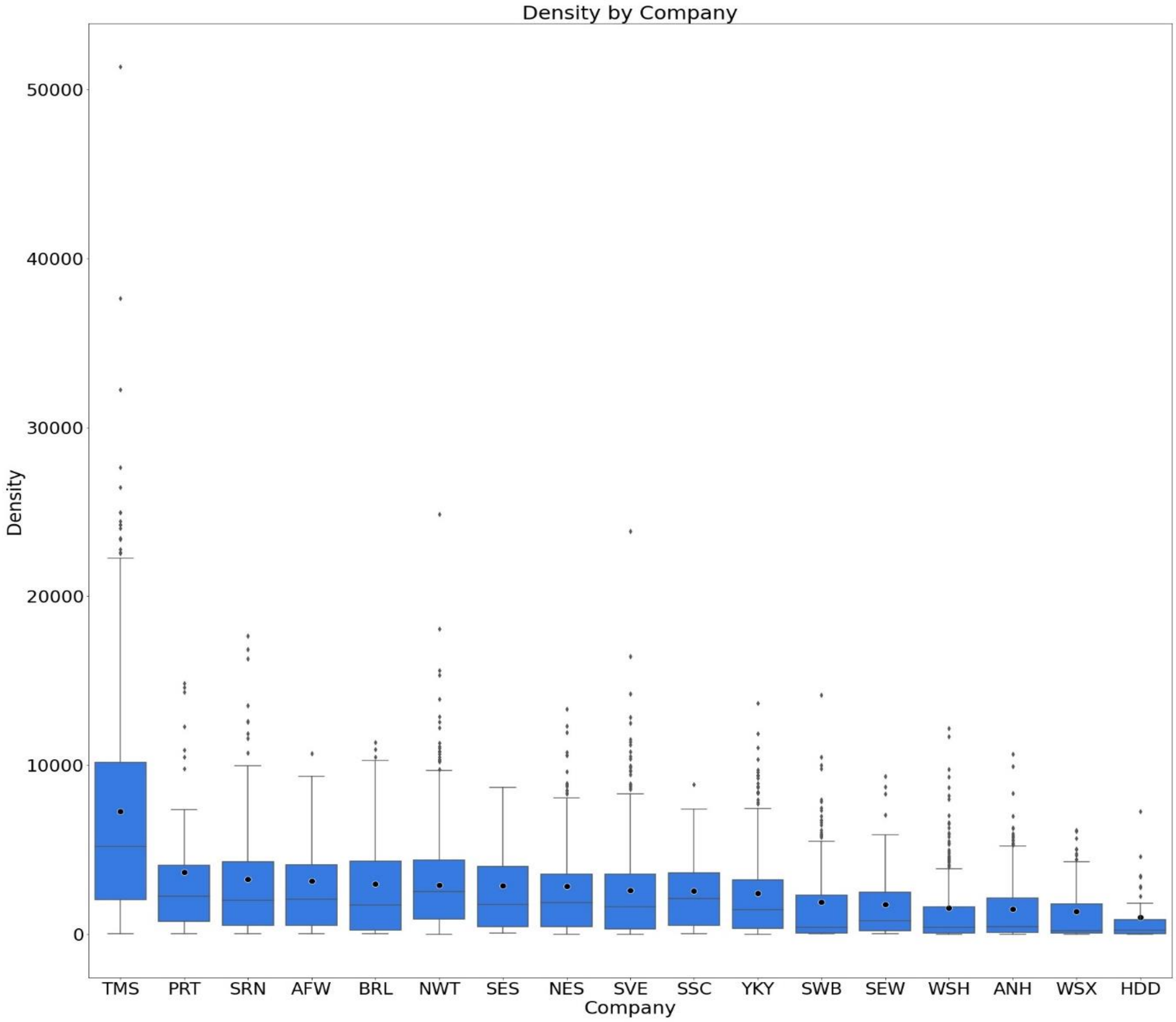


Figure 3.3.4 – Boxplot for companies' postcode sector population densities. The large black point for each company is the weighted average density. Created using 2011 census data.



Issues with considering density only at a company average level

The current weighted average measure produces a single density value per company. The squared term was added to generate a 'curve' of associated costs, acknowledging that there are costs associated with both sparsity and density. However, this assumes that companies have a uniform level of density and does not account for the challenges of having both sparse and dense areas within a company.

Another approach to accounting for density considered at PR19 was based on including separate metrics for the proportion of local authorities above or below a certain population density or sparsity threshold. This could be in addition to, or in place of, an average density metric. It starts to acknowledge the distribution of densities across a company in line with the way that engineering logic identifies that cost pressures will manifest.

However, density threshold data also suffers from issues around the degree of local authority disaggregation. Consequently, we have again used more granular postcode sector data to explore potential improvements. We believe that this is more likely to reveal more robust and appropriate thresholds.

Figure 3.3.5 shows a table of moments for each company based on the postcode sector distribution. A company with high kurtosis (the extent to which the data distribution has tails) suggests that there are a high number of density outliers. This is particularly problematic for companies like SVE, HDD, NWT, SWB and WSH. These company supply areas are likely to have a lot of rural areas alongside very dense outlier regions. However, these density outliers will not be accounted for by the weighted average density measure.

Skewness provides another view of this outlier issue – high skewness indicates that a company's postcode sectors are skewed heavily towards lower densities. Where high skewness is combined with high kurtosis, a company has to contend with mostly very rural areas, but with a significant number of very dense outliers as well.

Figure 3.3.5 – Table of the moments of the postcode sector density distribution.

Company	Kurtosis	Skewness	Standard Deviation	Mean
WSH	8.882	2.583	1756	1192
SVE	8.659	2.061	2449	2288
HDD	7.213	2.532	1522	890
SWB	5.723	2.072	2081	1513
NWT	5.577	1.531	2585	2994
SRN	4.936	1.873	2966	2839
PRT	3.545	1.871	3370	3162
TMS	3.441	1.325	5954	6699
SEW	3.009	1.56	1652	1491
ANH	3.005	1.586	1638	1319
YKY	2.677	1.484	2160	2113
NES	2.454	1.393	2279	2356
WSX	1.306	1.484	1466	1095
BRL	0.949	2.072	2753	2568
AFW	0.282	0.972	2348	2608
SSC	-0.164	0.576	1793	2254
SES	-0.326	0.84	2151	2378

Conclusion

We suggest removing the squared density term is a necessary starting point given it does not serve its intended purpose of reflecting cost drivers in areas of high density and sparsity.

From there, we suggest consideration of either:

- an explanatory factor that focuses on the extremes of density and sparsity, but based on more even and granular geographical designations (e.g. postcode sectors or DMAs) if it is possible to acquire that annual data for and each company within the dataset; or
- use of a symmetrical adjustment based on the true conditions that a company faces relative to the average density value. This could use the moment conditions displayed in *Figure 3.3.5*).

Water treatment complexity

There is a clear engineering expectation that water treatment costs should rise as the complexity of installed treatment processes increases (due to the external raw water quality pressures placed upon companies).

However, we have undertaken some analysis which suggests that there may be a case for changing the structure of the water treatment complexity band data that was used in PR19 models. This is because the existing data does not appear to be robustly describing the treatment complexity cost pressures seen across the sector.

Our current analysis is based on the statistical performance of changing model specifications using the existing data. However, we recognise that any amendments to the data or its use in models will require solid supporting engineering logic and evidence. Therefore, we have commissioned consultancy support to consider whether any of the treatment processes are placed in the wrong banding, and to explore whether having multiple treatment processes of the same band on the same site is significantly more costly than having just one treatment process in the same band.

Ensuring water treatment complexity bands are appropriate

We have undertaken analysis that challenges the statistical coherence of alternative specifications using the existing treatment complexity band data (i.e. the extent to which the treatment complexity data being used is actually sensitive to the treatment complexity costs incurred). ***Figure 3.3.6*** shows alternative model specifications that use alternative weighted average complexity definitions for the WRP2 specification. ***Figure A.3.1*** in ***Appendix A.3*** shows the same for the WW2 specification. The alternative definitions that we have tested are:

- *lnalt_wac1* – This makes no distinction between bands 2 and 3 or 4 and 5, i.e. no distinction between having one more expensive treatment process and 2 or more. i.e. weightings on bands 0-6 are 1, 2, 3, 3, 4, 4, 5.
- *lnalt_wac2* – As above, but with squared weightings. This represents treatment processes in bands 4 and 5 being far more expensive than treatments in bands 2 and 3. i.e. weightings on bands 0-6 are 1, 4, 9, 9, 16, 16, 25.
- *lnalt_wac3* – Original weighted average complexity breakdown but with squared weightings. i.e. weightings on bands 0-6 are 1, 4, 9, 16, 25, 36, 49.

Figure 3.3.6 – Regression coefficients (p-values) for WRP2 specifications with different definitions of weighted average treatment complexity. These use a panel of data from 2011-12 to 2019-20.

Variable	Original WRP2 Model	WRP2 alt_wac1	WRP2 alt_wac2	WRP2 alt_wac3
lnproperties	1.0195 (0.000)	1.0260 (0.000)	1.0309 (0.000)	1.0228 (0.000)
lnwac/lnalt_wac1/lnalt_wac2/lnalt_wac3	0.4712 (0.002)	0.6229 (0.000)	0.3462 (0.000)	0.2578 (0.047)
lnwedensitywater	-1.1681 (0.004)	-1.1311 (0.010)	-1.1221 (0.020)	-1.1513 (0.012)
lnwedensitywater2	0.0693 (0.016)	0.0667 (0.031)	0.0658 (0.053)	0.0680 (0.034)
constant	-6.1048 (0.000)	-6.3802 (0.000)	-6.5693 (0.053)	-6.2868 (0.034)
R^2	0.9186	0.9194	0.9156	0.9141
AIC	-9.32	-9.47	-8.54	-8.16
BIC	12.12	11.97	12.89	13.28
RESET	0.1884	0.2252	0.2402	0.1751
MAPE 2018-20	24.34%	24.12%	24.55%	24.88%

We would expect that if weighted average complexity (as it is currently defined) were a true reflection of treatment costs, it would be a significantly better predictor of WRP and WW costs than any alternative definition. However, reviewing the R^2, AIC, BIC, and MAPE metrics together, the quality of the WRP2 models change very little. Therefore, it is not possible to say that any one of them is significantly better than the others. The same is true for the WW2 models. This leads us to question the underlying predictive capability to the current data.

Accounting for the interaction of ground water and surface water complexities

Engineering logic would suggest that treatment cost pressures will differ depending on whether treatment assets are being supplied from ground or surface water resources. However, costs should also rise as complexity increases irrespective of the water resource used – a complex surface water treatment asset will incur more cost than a less complex one, as will a complex groundwater asset relative to a less complex one. A robust inclusion of treatment complexity in model specifications should account for both of these fundamental expectations.

The PR19 models consider treatment complexity irrespective of the water resource used. The use of the 3-6 complexity banding infers that all treatment assets greater than band 2 have an equivalent cost pressure. The distribution of treatment complexities across the industry shows that the vast majority of surface water sites have complexities greater than treatment band 3. Conversely, there is much more variation across the 3-6 boundary for groundwater sourced assets. This means that the treatment complexity band 3-6 coefficient currently used is solely describing cost pressures at groundwater sites, and that all surface water treatment has a similar cost profile. ***Figures 8.3.3-9*** are charts showing the proportion of various treatment complexity bandings and indicate that the variation in treatment bands 3-6 is almost entirely driven by groundwater.

We have modified the PR19 WRP model specification to account for surface water and groundwater complexities separately and also show the interaction of groundwater treatment cost relative to surface water treatment cost.

Figure 3.3.7 shows the regressions for the two updated WRP specifications. Unsurprisingly, the surface water band 3-6 category coefficient is insignificant/counterintuitive given that it there is very little differentiation between companies (almost all sites have complexities greater than band 3). But there is some explanatory power from including the proportion of volume treated at surface water relative to groundwater sites, indicating that collectively surface water sites have higher treatment costs than groundwater sites.

Therefore, we have repeated the analysis using treatment bands 4-6 (***Figure A.3.2*** in ***Appendix A.3***). This should create an appropriate distribution of surface water treatment complexities across the sector for the coefficient to differentiate between them. However, this updated specification continues to provide counter intuitive results for surface water complexity cost drivers (insignificant and negative implying the costs don't change as surface water complexity increases). Correlations between the SW and GW treatment drivers are low, so it is unlikely that the continuing modelling issues are a result of any multicollinearity. VIF test values of ~1.1-1.5 for all the additional variables used support this assumption.

Consequently, the analysis again suggests that the current surface water treatment complexity bands are not very effective at describing the true costs that are incurred at more complex surface water treatment sites.

Figure 3.3.7 – Regression coefficients (p-values) of the WRP specifications with GW and SW treatment split. These use a panel of data from 2011-12 to 2019-20.

Variable	WRP1 SW and GW Split	WRP2 SW and GW Split	WRP1 SW/GW with SW Control	WRP2 SW/GW with SW Control
lnproperties	1.0361 (0.000)	1.0157 (0.000)	1.0076 (0.000)	0.9951 (0.000)
pctwatertreated36_GW	0.0031 (0.026)	-	0.0030 (0.028)	-
pctwatertreated36_SW	0.0003 (0.866)	-	0.0017 (0.412)	-
lnwac_GW	-	0.2735 (0.012)	-	0.2343 (0.022)
lnwac_SW	-	-0.4636 (0.403)	-	-0.2334 (0.673)
pctwatertreatedSW	-	-	0.0043 (0.015)	0.0037 (0.038)
lnwedensitywater	-1.4251 (0.007)	-1.4796 (0.000)	-1.6509 (0.000)	-1.5831 (0.003)
lnwedensitywater2	0.0940 (0.016)	0.0940 (0.000)	0.1066 (0.000)	0.1024 (0.004)
constant	-3.9693 (0.097)	-3.9693 (0.016)	-4.3286 (0.002)	-3.9514 (0.115)
R^2	0.9185	0.9185	0.9269	0.9247
AIC	-7.11	-7.90	-11.49	-10.27
BIC	17.39	16.60	16.07	17.29
RESET	0.3061	0.4058	0.6428	0.7169

Conclusion

We have undertaken several analyses which suggest that the current approach to accounting for water treatment complexity should be reviewed as it doesn't support the expected engineering logic. The analyses all suggest that the current treatment complexity band data might be poorly defined such that they are not able to explain costs that are incurred at more complex sites.

Assuming the sensitivity of complexity data to cost pressures can be improved, it will then be important to include explanatory variables that describe the distribution of complexities seen in the industry at both groundwater and surface water sites.

We recommend a reassessment of the complexity bands. Our analysis suggests a potential for misspecification which could take one of multiple forms:

- Certain treatment processes are in the wrong band.
- Multiple 'higher cost' treatment processes are not significantly more expensive than a single 'higher cost' treatment process.
- Single 'higher cost' treatment processes are not significantly more expensive than multiple 'lower cost' treatment processes.

We have also commissioned consultancy support to study the engineering logic.

Water Resources cost drivers

We are conscious that the WRP specifications used at PR19 did not include any drivers that were specifically targeted at estimating water resources costs. This could mean that 'noise' is being added to the model because of water resources costs not being adequately captured.

We have commissioned consultancy support to consider what might be driving water resources costs. This should reinforce the prior engineering expectations that in turn should govern the specification of models.

Identifying and testing appropriate remedies is hampered by the fact that very few commonly reported variables are available that might explain the variation in the water resources costs. A possible solution is to include the proportion of various sources. We have explored potential model specifications including the proportion of distribution input extracted from rivers and proportion of distribution extracted from boreholes. Regressions including these variables are included in ***Figure 3.3.8***.

Figure 3.3.8 – Regression coefficients (p-values) from various WRP and WR + RWD specifications, using a panel of data from 2011-12 to 2019-20.

Variable	WRP1	WRP2	WRP1 Plus WR Drivers	WRP2 Plus WR Drivers	WR + RWD 1	WR + RWD 2
lnproperties	1.0246 (0.000)	1.0195 (0.000)	0.9857 (0.000)	0.9737 (0.000)	0.9806 (0.000)	0.9745 (0.000)
pctwatertreated36	0.0079 (0.000)	-	0.0059 (0.004)	-	0.0028 (0.377)	-
lnwac	-	0.4712 (0.002)	-	0.2564 (0.244)	-	0.1980 (0.494)
lnwedensitywater	-1.6569 (0.000)	-1.1681 (0.004)	-0.9773 (0.141)	-0.7019 (0.305)	-0.2483 (0.747)	-0.0770 (0.913)
lnwedensitywater2	0.1046 (0.000)	0.0693 (0.016)	0.0582 (0.191)	0.0391 (0.392)	0.0150 (0.781)	0.0028 (0.955)
prop_rivers	-	-	-0.4594 (0.041)	-0.4770 (0.039)	-0.5306 (0.014)	-0.5269 (0.012)
prop_boreholes	-	-	-0.4680 (0.018)	-0.6205 (0.002)	-0.5741 (0.006)	-0.6121 (0.001)
constant	-4.4666 (0.001)	-6.1048 (0.000)	-5.9087 (0.010)	-6.5498 (0.008)	-9.8173 (0.000)	-10.3785 (0.000)
R^2	0.9308	0.9186	0.9253	0.9159	0.9348	0.9351
AIC	-16.17	-9.32	-18.53	-14.58	19.13	19.40
BIC	5.26	12.12	9.03	12.99	46.69	46.96
RESET	0.4995	0.1884	0.5599	0.4598	0.4011	0.3235
MAPE 2018-20	21.96%	24.34%	20.52%	21.64%	19.96%	19.92%

While these expanded models do not appear to be a significantly better fit to the data than the original WRP specifications, we believe it is important to include Water Resources (WR) cost drivers for the same reasons as we mentioned in ***Chapter 3.1*** concerning growth. ***Figure A.4.1*** in ***Appendix A.4*** shows efficiencies changing systematically with the removal of WR and Raw Water Distribution (RWD) costs in proportion of WRP costs that are made up by WR and RWD. This would not be expected if water resources costs were analogous to Treatment costs and sensitive to the existing model coefficients.

Further, we can see from ***Figure 3.3.9*** that the variables remain significant when Water Treatment costs are used as the dependent variable. But (with the exception of scale and the constant) become insignificant when just WR and RWD costs are used as the dependent variable. This again suggests that while the existing cost drivers can adequately describe the distribution of treatment costs, they cannot not do the same for water resources costs. A similar analysis with the expanded WR + RWD specifications in ***Figure 3.3.8*** shows that the water resources drivers (proportion of distribution input from rivers and boreholes) remain significant. As with growth, this suggests there is nothing to account for the variation in the WR and RWD portion of WRP costs. Instead, water resources costs are treated as equivalent for each company relative to scale, and therefore any variation in proportion is treated as (in)efficiency. The addition of the water resources drivers would help to remedy this issue.

Figure 3.3.9 – Regression coefficients (p-values) from the original PR19 WRP specifications, using WT and WR + RWD as the dependent variables and using a panel of data from 2011-12 to 2019-20.

Variable	WT1	WT2	WR + RWD 1	WR + RWD 2
lnproperties	1.0306 (0.000)	1.0255 (0.000)	1.0304 (0.000)	1.0231 (0.000)
pctwatertreated36	0.0091 (0.000)	-	0.0047 (0.131)	-
lnwac	-	0.5227 (0.022)	-	0.3775 (0.221)
lnwedensitywater	-1.9799 (0.000)	-1.4616 (0.004)	-1.0525 (0.152)	-0.7180 (0.331)
lnwedensitywater2	0.1241 (0.000)	0.0868 (0.015)	0.0698 (0.176)	0.0457 (0.372)
constant	-3.6875 (0.036)	-5.4046 (0.003)	-8.1158 (0.001)	-9.3338 (0.000)
R^2	0.8927	0.8746	0.9171	0.9173
AIC	29.44	35.69	26.00	27.24
BIC	50.87	57.12	47.44	48.68
RESET	0.7078	0.3773	0.0729	0.0225
MAPE 2018-20	29.12%	32.23%	21.79%	22.35%

Conclusion

We recommend including water resources cost drivers in the WRP specification. These might be taken from the current pool of variables or may require new variables to be collected. Not doing so effectively treats any variation in water resources costs as noise.

Water Network complexity

There has been a lot of debate as to whether Average Pumping Head (APH) or Boosters per Length (BPL) is the correct variable to account for water network complexity. We believe that there is merit in including both variables in the TWD and WW model specifications. Fundamentally, they represent two different aspects of cost.

We have commissioned consultancy support to consider whether the engineering logic outlined below is correct, and whether there is any further logic to back up the claim that both APH and BPL should be included in the regressions.

The engineering logic for considering APH and BPL as separate cost drivers

APH increases as the requirement to pump water within the network increases. This is largely driven by the topography of the supply area and might be because of very flat or very hilly regions. A single booster pumping station with a high lift and capacity could be equivalent to multiple smaller pumping stations delivering an equivalent volume and lift. Both scenarios would incur significant energy costs.

BPL increases with the number of booster pumping stations in a given area. There may be many low capacity / lift booster pumping stations in a particular area of network (for example, in supply areas with small populations in hilly areas). Where booster pumping stations are required, they are usually matched with corresponding pressure release valves. Simplistically, a booster lifts the water on one side of a hill, then the pressure release valve is required on the other side to manage pressure and velocity and thus bursts elsewhere in the network.

Where boosters lift relatively small volumes of water, they will not individually contribute much to the aggregated APH value. However, these additional assets and corresponding pressure release valves all require ongoing capital maintenance and create weak points in the network that are prone to leaks and bursts and hence create further capital maintenance pressures. This therefore creates a capital maintenance cost pressure.

In short, APH is likely to act as a proxy for Opex (power) costs, while BPL is more likely to reflect long term capital maintenance and asset heath cost pressures within the network.

The statistical case for including APH and BPL as separate model explanatory factors

Figure 3.3.10 – A scatter graph showing the relationship between boosters per length and Average Pumping Head in Treated Water

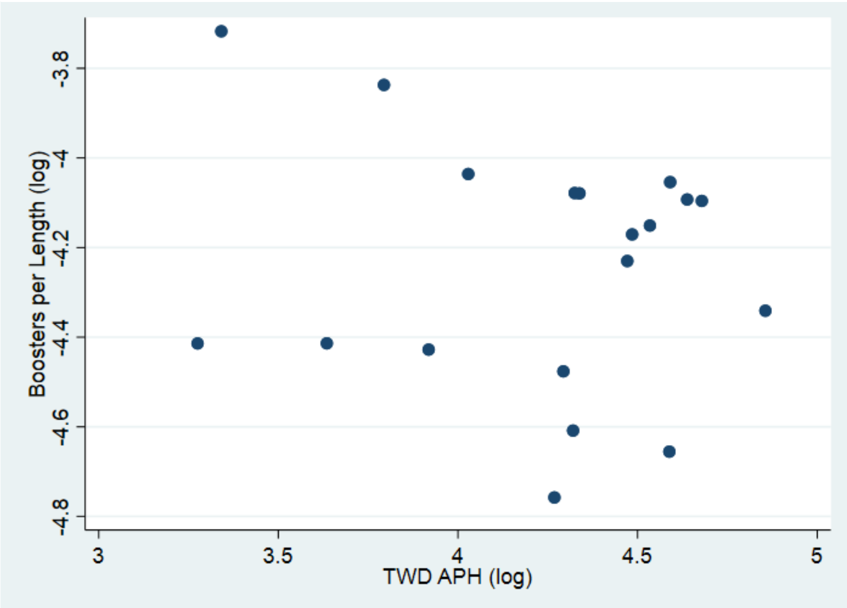


Figure 3.3.10 shows a scatter chart of mean log BPL and mean log APH in Treated Water Distribution. From a visual inspection, we can see that there is very little correlation between the two variables, and certainly not the strong correlation we would expect if they were indicative of the same cost pressures. The Pearson correlation coefficient between them is -0.0898. This is far from a strong correlation, consequently there should not be a statistical problem with including them both in the same regression.

For the TWD specification shown in ***Figure 3.3.11***, BPL and APH are both significant, while APH is insignificant in the WW specifications, shown in ***Figure A.2.1*** and ***Figure A.2.2*** in ***Appendix A.2***. Considering the R^2, AIC, BIC, and MAPE metrics together, the joint specifications appear to perform better statistically, and the coefficients only change marginally between the joint specifications and the single network complexity variable specifications which is indicative of robustness. As a result, we take this as strong evidence for the inclusion of both variables in the TWD and WW specifications. However, we acknowledge that APH data is currently known to be subject to inconsistencies which Ofwat is seeking to address.

Figure 3.3.11 – Regression coefficients (p-values) with just BPL, APH + BPL, just APH for the TWD specification. These use a panel of data from 2011-12 to 2019-20.

Variable	Original TWD Model	TWD with APH + BPL	TWD with just APH
lnwedensitywater	-3.1969 (0.000)	-3.4112 (0.000)	-4.1374 (0.000)
lnwedensitywater2	0.2562 (0.000)	0.2696 (0.000)	0.3117 (0.000)
lnlengthsofmain	1.0620 (0.000)	1.0696 (0.000)	1.0916 (0.000)
lnboostersperlength	0.5504 (0.000)	0.4988 (0.000)	-
lnaph_twd	-	00994 (0.047)	0.1117 (0.042)
constant	6.1092 (0.000)	6.2366 (0.000)	6.8898 (0.000)
R^2	0.9623	0.9652	0.9580
AIC	-100.87	-104.34	-97.46
BIC	-79.43	-79.84	-76.02
RESET	0.0481	0.0403	0.0408
MAPE 2018-20	20.45%	19.39%	20.42%

Conclusion

We recommend including both BPL and APH in the TWD and WW specifications at PR24 rather than only considering one of these variables. Although they can both be considered Network Complexity drivers, they describe different elements of cost and statistically they perform well together in the models.

Wastewater Treatment economies of scale

Anglian Water have challenged the way in which economies of scale at sewage works are specified in the PR19 models. They suggest that the current metric does not reflect ongoing economies of scale at larger works and that there isn't enough variation across the sector using the current size bands.

The engineering logic for increasing economies of scale at large works is uncontentious. However, the selection of a threshold should also be based on engineering expectations. Consequently, an argument based on sector marginal costs feels less compelling. It has the potential to be biased by inefficient companies pushing up the marginal cost for a particular band, particularly where there are relatively few sewage works in that band. There may also be different treatment processes at these works which are not accounted for in the analysis but which could substantially change the presented marginal costs. We also note that using disaggregated size band data above band 6 could reduce the number of years for which accurate data is available.

We consider that the marginal cost analysis done by Anglian does not give a clear basis for selecting a threshold of band 8 and above as proposed. The marginal cost difference between band 8 and Ofwat's band 6 is negligible, as is the marginal cost difference between band 7 and band 8. The analysis also suggests that a diseconomy of scale may exist in that the frontier costs in bands 9 and 10 are higher than in bands 7 and 8.

Conclusion

We are not opposed to reconsidering the economies of scale bandings for wastewater treatment works. However, there needs to be solid engineering rationale behind the selection of thresholds linked with material changes in incurred marginal cost. This is to avoid any potential of weakening the model.

Accounting for legitimate changes in cost over time

We believe it is essential to account for changes in base costs over time, whether as a result of changes in quality that are not included in the models or exogenous changes in the conditions in which companies operate (e.g. raw water deterioration and climate change).

We understand that Ofwat has concerns with the addition of time variables to the models, and has suggested that there might be interactions with the efficiency challenge. However, it might be more appropriate to account for issues with time through the model rather than through the efficiency challenge. The addition of time variables to the model is likely to account for all the time related issues. Real Price Effects are also an external driver of cost outside of company control, which increase over time. If time effects are included in the models, this could open up a discussion about whether the RPE adjustment to the frontier shift assumptions may then be removed from the efficiency challenge calculation.

There are two ways that we consider this could be done.

- The inclusion of **indicator variables** which could be used to account for:
  - AMPs;
  - the years within an AMP; or
  - individual years that we consider to be 'exceptional' and may have induced costs that are outside the norm.
- The inclusion of a **time trend**.

Indicators

AMP indicators account for exogenous change between AMPs. We might consider these to be accounting for step-changes in quality across AMP periods. When well specified, indicators help to remove noise from the model and therefore increase their predictive power. Indicator coefficients are interpreted relative to data that has not been given an indicator. We would therefore not allocate an indicator to the most recent AMP. This assumes that the future costs will be most applicable to recent costs relative to earlier costs.

AMP Year indicators account for AMP investment profiles that are common across the sector. We generally see lower levels of spending in the first years of an AMP and higher spending towards the end. AMP Year indicators can take account of this. They might also serve to highlight companies that are in a ‘trough’ if their expenditure profile moves in opposition to the predicted values which would consider the profiles of the industry as a whole.

Year indicators should be used to account for any years where costs could be considered exceptional and are unlikely to be repeatable. The pandemic years would be a good candidate for inclusion here. Increased expenditure on PPE and atypical bad debt provision are two examples of increased expenditure that may not be seen again, certainly not going forward into the next AMP.

Figure 3.3.14 shows the WW1 specification from PR19, with variations adding in an indicator for AMP5, then adding AMP Year indicators, then adding an indicator for 2019-20 accounting for covid concerns late on in the financial year. We can see that the main coefficients of interest remain significant throughout, that the RESET test is always passed, the R^2 increases, and the AIC and BIC become more negative, indicating there is evidence for the final model providing the best fit.

Of course, there are issues with these specifications. For example, we only have one year of ‘ampyear1’ data since there is no 2010-11 data in the dataset, and so it’s essentially acting as a 2015-16 indicator. These serve as an example of how these models can be specified and show that the models still work when these additional variables are included. Given the regulatory and legislative changes that occur between AMPs and the subsequent impact on operating costs (e.g. future operation of enhancement assets), it seems sensible to include the AMP indicators at the very least.

We can also consider interaction terms to account for any structural breaks in the data if we believe that the relationship between the explanatory variables and the dependent variable change between AMPs. An output and accompanying explanation for this can be seen in ***Figure A.5.1*** in ***Appendix A.5***.

Figure 3.3.14 – Regression coefficients (p-values) from the original PR19 WW1 specification, with the addition of an AMP5 indicator, the iterative addition of AMP Year indicators and the iterative addition of a 2019-20 indicator. These use a panel of data

Variable	Original WW1 Model	WW1 + AMP5	WW1 + AMP5 + AMP Years	WW1 + AMP5 + AMP Year + 2019-20
lnproperties	1.0365 (0.000)	1.0343 (0.000)	1.0306 (0.000)	1.0304 (0.000)
pctwatertreated36	0.0055 (0.000)	0.0046 (0.000)	0.0034 (0.000)	0.0033 (0.000)
lnwedensitywater	-2.3239 (0.000)	-2.2055 (0.000)	-2.0506 (0.000)	-2.0418 (0.000)
lnwedensitywater2	0.1652 (0.000)	0.1561 (0.000)	0.1446 (0.000)	0.1441 (0.000)
lnboostersperlength	0.3531 (0.005)	0.3308 (0.009)	0.3053 (0.015)	0.3068 (0.015)
amp5	-	-0.0458 (0.134)	-0.0972 (0.002)	-0.0626 (0.051)
ampyear1	-	-	-0.1903 (0.000)	-0.1037 (0.000)
ampyear2	-	-	-0.0204 (0.383)	0.0486 (0.187)
ampyear3	-	-	0.0131 (0.618)	0.0820 (0.040)
ampyear4	-	-	-0.0072 (0.534)	0.0628 (0.016)
year2020	-	-	-	0.1440 (0.000)
constant	-0.3734 (0.777)	-0.7218 (0.554)	-1.1255 (0.354)	-1.2374 (0.298)
R^2	0.9732	0.9732	0.9746	0.9754
AIC	-139.11	-140.89	-164.42	-172.88
BIC	-114.61	-107.34	-124.61	-130.00
RESET	0.1276	0.0974	0.1848	0.6697

Time Trend

Time trends essentially serve as a control, accounting for potentially a wide variety of factors that aren't included in the model. For example these might be:

- continuous improvements in the quality of service;
- variables that haven't been included in the model that change year on year and are associated with changes in cost; or
- the effects of changes in real prices.

There is no reason that the trend component should be positive, for example it could be accounting for changes in technology that improve efficiency, resulting in a negative coefficient.

A brief description of time trends in the regression context is shown in ***Appendix A.6***.

At PR19, the effect of rising input prices from labour costs over and above inflation was netted off the frontier shift efficiency challenge (the real price effect (RPE) adjustment). This effectively acknowledged that future costs are likely to increase with time that are outside of company control and therefore cannot be managed by productivity improvements. The inclusion of a time trend could similarly account for these input price pressures where they are already exhibited in the historic data set. However, it would also account for any changes in technology or complexity that might alter the wider cost base over time which also cannot be managed by productivity gains.

Therefore, we consider a time trend which accounts for all costs should be used to account for a more general group of time dependent cost pressures. It would then be sensible to consider whether the inclusion of both would lead to an unnecessary double protection against input price risks.

A set of regressions showing the addition of a time trend to the PR19 water models is shown in ***Figure 3.3.15*** with the tests for model quality for the original models shown alongside the tests for the time trend models. We can see that, for the TWD and WW specifications, the time trend is significant, and by AIC and BIC provides a significantly better fit to the data.

For WRP models, the time trend component appears insignificant, but the models do not appear to be significantly worse than their counterparts without the time trend. The variable could be excluded as a strict interpretation would be that costs do not change as consistently over time in this model. However, the overriding logic of the time trend acting as a model control remains.

Figure 3.3.15 – Regression coefficients (p-values) from PR19 specifications with the addition of a time trend, using a panel of data from 2011-12 to 2019-20.

Variable	WRP1	WRP2	TWD	WW1	WW2
lnproperties	1.0223 (0.000)	1.0184 (0.000)	-	1.0319 (0.000)	1.0252 (0.000)
pctwatertreated36	0.0071 (0.000)	-	-	0.0038 (0.000)	-
lnwac	-	0.3471 (0.019)	-	-	0.3653 (0.000)
lnwedensitywater	-1.5257 (0.000)	-1.0646 (0.026)	-2.8760 (0.000)	-2.0853 (0.000)	-1.8163 (0.000)
lnwedensitywater2	0.0951 (0.000)	0.0618 (0.064)	0.2303 (0.000)	0.1471 (0.000)	0.1278 (0.000)
lnlengthsofmain	-	-	1.0510 (0.000)	-	-
lnboostersperlength	-	-	0.4640 (0.001)	0.3117 (0.010)	0.3197 (0.005)
timetrend	0.0067 (0.500)	0.0091 (0.344)	0.0248 (0.000)	0.0157 (0.000)	0.0146 (0.002)
constant	-4.8384 (0.000)	-6.2932 (0.000)	4.7705 (0.001)	-1.1830 (0.317)	-2.2144 (0.042)
R^2 (R^2 no trend)	0.9293 (0.9308)	0.9151 (0.9186)	0.9668 (0.9623)	0.9733 (0.9732)	0.9741 (0.9750)
AIC (AIC no trend)	-15.09 (-16.18)	-9.02 (-9.31)	-128.24 (-100.87)	-148.86 (-139.11)	-147.70 (-140.53)
BIC (BIC no trend)	9.41 (5.26)	15.48 (12.12)	-103.74 (-79.43)	-121.30 (-114.61)	-120.13 (-116.03)
RESET (RESET no trend)	0.5648 (0.4995)	0.3469 (0.1884)	0.1695 (0.0481)	0.1539 (0.1276)	0.0530 (0.0434)
MAPE 2018-20 (MAPE no trend)	22.07% (21.96%)	24.37% (24.34%)	19.47% (20.45%)	15.00% (14.98%)	15.21% (14.77%)

There is further discussion concerning the use of a time trend in ***Chapter 5***.

Conclusion

We should consider the addition of variables that control for changes in cost over time that are inadequately captured by the other variables in the model including the mitigation of environmental concerns and ongoing work following improvements in quality. Time variables are an important control, and in the long run their omission could lead to misspecification problems.

If time variables are included, there could be a discussion around the removal of the RPE component of the efficiency challenge to avoid any risk of customers paying twice for price increases that influence the coefficients on these time variables.

3.4 Sample period selection

Summary of content / messaging in Ofwat's consultation document

- Ofwat questions whether it is appropriate to consider the full time series available to model base costs given there may be structural breaks in the data, although its working assumption is to utilise the full data series available (back to 2011-12).
- Ofwat considers that it might be appropriate to use business plan forecasts in the base econometric models to account for future cost pressures.

Consultation Questions asked
Sample period selection
1. Do you agree that we should utilise the full historical data series available to develop the wholesale base cost models at PR24 (from 2011-12 onwards) unless there is clear justification for using a reduced time series (e.g. structural break that cannot be addressed through other remedies)?
2. Should we consider including business plan forecasts in our wholesale base cost models at PR24?

Severn Trent response

- Compared to the wider uses of econometric modelling, the data panel being used here is very small. Therefore, we consider the longest possible data panel should be used (i.e. back to 2011-12 where robust data allows). Time indicators or time trends should then be used to remedy any potential issues from using the full length of the data set.
- Our view is that business plan forecasts should not be used in the regressions as they are likely to materially increase uncertainty and reduce the quality of models. This will impact the accuracy of the coefficients and the efficiency challenge.
- A forward-looking efficiency challenge is a pragmatic way to consider future cost pressures, subject to caveats on the implementation of the challenge.

The robustness and performance of econometric models is strongly influenced by the size of the dataset that is being used to construct them. Compared to other typical uses of econometric models where thousands of observations are standard, the size of the dataset being used is very small. Therefore, we strongly support the use of the full data panel back to 2011-12.

However, as the length of the panel increases, so too does the potential for variations across the duration of the panel. This can be due to year-to-year fluctuations or more gradual trends over time. Consequently, we believe that there is an increasing justification to account for the effects of time across the panel when specifying the model. This would mean that the impacts of atypical years (e.g. the impacts of the pandemic or particularly adverse weather) and any trends in external pressures and efficiency performance improvements in the historic data set will be recognised and their impact on forecasts minimised. This has been discussed in detail in ***Chapter 3.3*** of this response.

We strongly counsel against the use of business plan forecasts to develop models. This is likely to add significant uncertainty and will have significant endogeneity issues. The fundamental concern is that business plan values are not actual values. They are subject to inherent uncertainty, contrasting forecasting capability and differing levels of risk assumed by companies with respect to opportunities for future efficiencies, anticipated deterioration rates and external pressures that will be faced. These uncertainties will inherently increase model noise. These concerns are visualised in ***Figure 3.4.1*** which show the absolute variation between forecast and actual data in 2018-19 and 2019-20. It shows for three example model dependent variables and two example explanatory factors that variations between actual and forecast values can be very material and sporadic. The Mean Absolute Prediction Errors of these examples are also displayed in ***Figure 3.4.2***.

Figure 3.4.1 – Absolute prediction error percentage of actual vs business plan values for the 2018-19 and 2019-20 financial years.

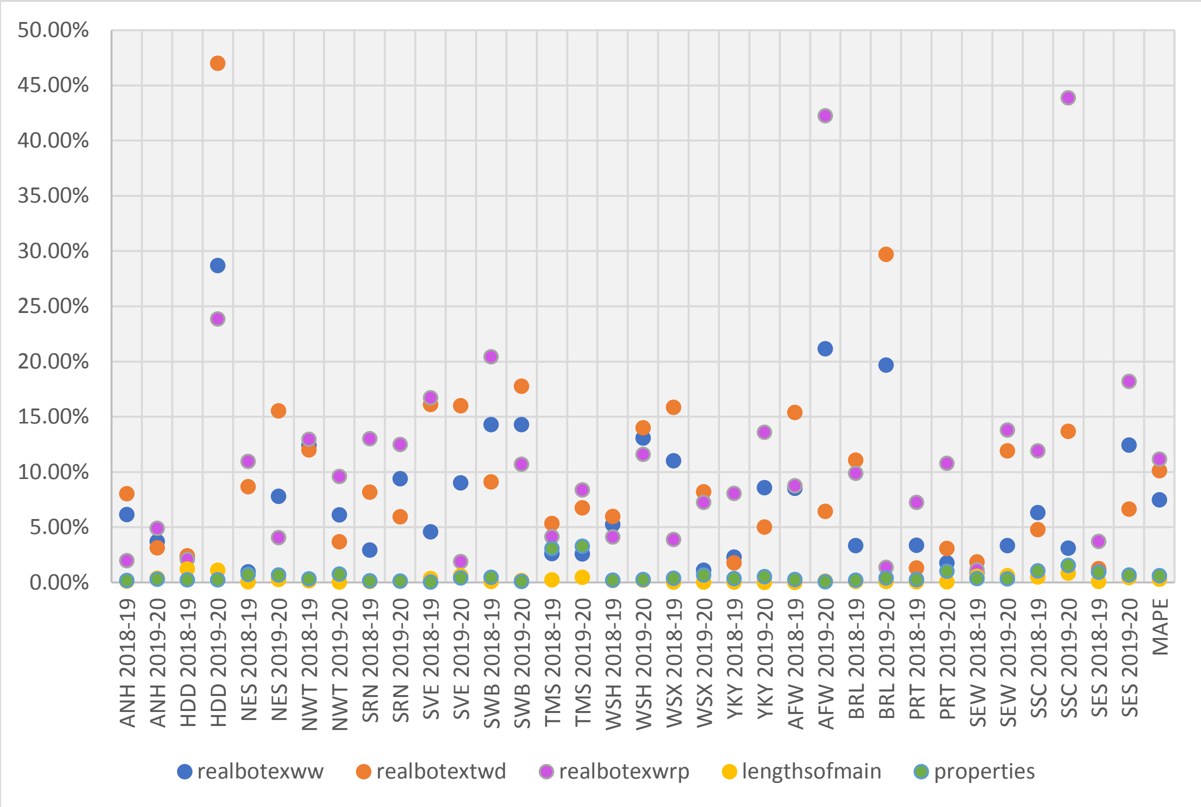


Figure 3.4.2 – Mean absolute prediction errors of actual vs business plan values for all companies for the 2018-19 and 2019-20 financial years.

realbotexww	realbotextwd	realbotexwrp	lengthsofmain	properties
7.46%	10.09%	11.15%	0.27%	0.59%

In the analysis below, we show that business plan forecasts can lead to problems with respect to model performance and the significance / coherence of the coefficients. This is in violation of the 4th cost assessment principle – robust econometric models.

Testing the impact of Business Plan Forecasts in Base Cost Regressions

We do not agree with the use of business plan forecasts in the base cost regressions and believe this would undermine the statistical validity of the models, adding uncertainty and reducing the robustness of the coefficients.

Figure 3.4.3 shows the change in coefficients for one of the wholesale water specifications used at PR19. The coefficients change substantially when business plan forecast data is added. In the first re-specification, just the financial years 2018-19 and 2019-20 are business plan forecasts, with the remaining years from 2011-12 to 2017-18 being actuals. Further, the coefficient on boosters per length reduces in significance in the business plan regressions, which could be used as an argument for its exclusion.

When the full AMP7 business plan values are used (2020-21 to 2024-25, with 2011-2020 as actuals), the models break down further, with very large changes in coefficients and further reductions in the significance of variables. While this isn't an issue in and of itself and would not be contentious where actual values are used – when considered alongside the concerns with the forecasts themselves, this becomes problematic.

Figure 3.4.3 – A table showing water regression coefficients (p-values) when actual values are used, and when business plan values are used. Changes shown are relative to the 'Actuals 2018-20' results.

Variable	Actuals 2018-20	Business Plan 2018-20	% Change in Coeff	Business Plan 2020-25	% Change in Coeff
lnproperties	1.0364 (0.000)	1.0439 (0.000)	0.72%	1.0306 (0.000)	-0.56%
pctwatertreated36	0.0055 (0.000)	0.0052 (0.000)	-5.45%	0.0027 (0.000)	-50.91%
lnwedensitywater	-2.3239 (0.000)	-2.3159 (0.000)	-0.34%	-1.8039 (0.000)	22.11%
lnwedensitywater2	0.1652 (0.000)	0.1617 (0.000)	-2.11%	0.1245 (0.000)	-24.45%
lnboosterperlength	0.3531 (0.005)	0.1781 (0.076)	-49.56%	0.1888 (0.094)	-54.38%
constant	-0.3734 (0.777)	-1.0780 (0.360)	-188.70%	-2.4160 (0.075)	-547.03%
R^2	0.9732	0.9752	-	0.9654	-
RESET	0.1276	0.4464	-	0.6029	-
Efficiency Challenge (All Models)	0.9456	0.9583	1.34%	0.9964	5.37%
Sector-Wide Totex Allowance (All Models)	£25,186m	£25,152m	-0.13%	£24,433m	-2.99%

Figure 3.4.4 shows the equivalent changes in coefficients for a sewage collection specification from PR19. It shows similar changes in the coefficients and p-values. Interestingly, this regression shows that the coefficient on *lnwedensitywastewater* becomes significant at the 10% level in the 'Business Plan 2018-20' regression. This illustrates the risk that models and variables might be accepted on statistical grounds despite the fact that the underlying data robustness has reduced.

In both the water and waste regression tables, the RESET test suggests that we do not expect there is misspecification in any case but the 'Business Plan 2018-20' SWC specification. The R^2 values also change little. Again, this highlights that we should be cautious not to select a model simply because these tests are satisfied.

Figure 3.4.4 – A table showing wastewater regression coefficients (p-values) when actual values are used, and when business plan values are used. Changes shown are relative to the 'Actuals 2018-20' results.

Variable	Actuals 2018-20	Business Plan 2018-20	% Change in Coeff	Business Plan 2020-25	% Change in Coeff
lnsewerlength	0.8649 (0.000)	0.8422 (0.000)	-2.62%	0.7761 (0.000)	-10.27%
lnpumpingcapperlength	0.5144 (0.019)	0.4010 (0.014)	-22.05%	0.1161 (0.476)	-77.43%
lnwedensitywastewater	0.1875 (0.125)	0.1801 (0.098)	-3.95%	0.0915 (0.325)	-51.20%
constant	-6.1084 (0.000)	-5.7653 (0.000)	5.95%	-4.3125 (0.000)	29.40%
R^2	0.8925	0.9012	-	0.8800	-
RESET	0.1069	0.0429	-	0.3038	-
Efficiency Challenge (All Models)	0.9731	0.9756	0.26%	0.9845	1.17%
Sector-Wide Totex Allowance (All Models)	£22,654m	£22,659m	0.02%	£22,028m	-2.76%

Forward-Looking Efficiency Challenge

We believe that the use of a forward-looking efficiency challenge would be a pragmatic way to account for future challenges. However, given the uncertainty around forecast variables, the forecast costs will necessarily be less robust than the estimated backward-looking costs. As a result, the forward-looking efficiency challenge should:

- be set at a less stringent level than the backward-looking efficiency challenge (given the challenge should be related to model quality); and
- account for any short-termist or overly risky business plans by removing unsuitable data points from benchmark selection. This is outlined in more detail in ***Chapter 5*** of our response.

Conclusion

Including forward looking costs in base cost models will materially reduce their quality and predictive capability. Therefore, we do not consider it to be a viable approach.

However, there may be merit in using a carefully applied forward looking efficiency challenge to help account for future pressures not seen in the historical dataset.

4. Cost adjustment claims

Summary of content / messaging in Ofwat's consultation document

- Ofwat plan to retain the cost adjustment process for PR24.
- Ofwat proposes that base and enhancement claims should be treated separately at PR24.
- Ofwat considers that claims should be symmetrical by default and invites companies to evidence the base cost impact for other companies when claims are submitted.
- Ofwat suggests that claims that were rejected at PR19 will not be reconsidered at PR24 unless there has been a material change in the company's circumstances, there is a change in the evidence available, or the cost drivers in the base models at PR24 change.
- Ofwat will maintain the requirement for a high standard of evidence and retain the use of materiality thresholds and the gated assessment to assess cost adjustment claims at PR24.
- Ofwat will consider providing additional guidance on the estimation of the implicit allowance.

Consultation Questions asked
14. Do you agree that the cost adjustment claim process at PR24 should be separated between base (wholesale and residential retail) and enhancement claims?
15. What base cost adjustment claims (wholesale and residential retail) would you consider submitting if the PR19 cost models were used to assess efficient costs at PR24?
16. What additional cross-sector data should be collected to support the submission of the claims indicated in response to the previous question? Please describe and explain the rationale behind the additional data that you consider should be collected and provide a draft definition.
17. How can the cost adjustment claim guidance be enhanced to improve the quality of cost adjustment claim submissions?
18. Would an early cost adjustment claim submission be welcome at PR24?

Severn Trent Response

We welcome the commitment to retain a cost adjustment process. This is a critical 'safety valve' to the cost adjustment process. It facilitates the principle of seeking sensibly simple econometric base models which will not be able to capture all drivers of cost that affect companies. However, for the adjustment process to be effective and manage the inherent risk associated with sensibly simple models, it must be administered in a way that allows companies to apply for adjustments and have a reasonable expectation of success where a proportionate case is put forward.

Defining base and enhancement claims

It is sensible to consider adjustment claims for base and enhancement expenditure as distinct from each other. This will enable the assessment of claims more appropriately tailored to the different circumstances of base and enhancement expenditure. We would be happy to continue to engage on the development of more targeted criteria for both base and enhancement claims through the work of the CAWG.

The case for adjustments and consideration of materiality

We understand the premise of symmetrical adjustments where they relate to model specification and think that they can play a key role in improving the wider cost assessment approach (principle 6). This can be a critical way of improving the coverage of cost assessment where long data series are not available for inclusion in models. As the scope of the econometric models account for the vast majority of base expenditure, it stands to reason that symmetrical adjustments should be applied where material cost drivers are not covered by explanatory variables. In these cases, models will not appropriately allocate these costs between companies. This is in line with the PR19 growth post modelling adjustment.

However, there is a need for non-symmetrical adjustments.

- These will most likely be where the costs relate to one company only (i.e. any negative cost adjustment would be trivial or zero). Therefore, we would urge that the application of all symmetrical counter adjustments are only applied if they satisfy a materiality threshold. This would make to process more manageable and allow for situations where symmetrical adjustments are not appropriate.
- They may also be required for costs which have not been incurred by the sector in the past and hence cannot be predicted by econometric models. In these cases, no costs are being distributed between companies that would be subject to a counter symmetrical adjustment. We discuss the case for such a claim relating to the future operation of sewage works with tight P-consents below.

We note that Ofwat are stressing the importance of atypicality when making cost adjustment claims. Whilst we agree that it is an important characteristic of non-symmetrical claims, this is not necessarily the case for symmetrical claims. We think that symmetrical adjustments should be considered more as a tool to account for the impact of cost drivers on all companies where it is not possible or desirable to take account of them within the econometric models.

The interaction of symmetrical adjustments and econometric model specification

A key attraction of potentially using symmetrical adjustment claims to address some of the model specification issues identified are the impacts on data requirements and allowing econometric models to remain sensibly simple. We are fully aware of the difficulty of developing models that satisfy Ofwat's cost assessment principles and adequately reflect the individual requirements of a wide range of companies that have many different pressures and circumstances.

A particular challenge will be the sourcing of data that can be used in the modelling panel (i.e. back to 2011). This means that explanatory factors are either limited to information that is already collected or new data that would likely require substantial back casting and iterations of data collection to ensure appropriate levels of consistency. Both situations are likely to impact on the ability to identify robust modelling solutions.

Additionally, we are aware that adding too many additional cost drivers into model specifications could start to conflict with the desire for sensibly simple econometric models. This could signal a move back towards the more highly specified 'full' totex models used at PR14 which were subjected to challenge by the CMA in the Bristol Water appeal. Adding in additional explanatory drivers is likely to increase the risk of overfitting (as discussed in ***Chapter 2***) and its associated impact on predictive performance.

Applicability for making adjustment claims

We do not consider it would be appropriate to systematically reject potential cost adjustment claims that have not previously been raised or have been rejected at prior reviews. Understanding and interpreting econometric information is complex, and additional information/knowledge may have come to light since PR19. Therefore, it does not feel appropriate to hold companies to historical (potentially out of date) understanding and decisions previously made. In accepting our fast-track status at PR19 we needed to consider the proposed price review settlement in the round rather than respond to all potential areas of concern.

Implicit allowance

The implicit allowance (IA) sets out the extent to which cost adjustment claims are partially (or fully) accounted for within Ofwat's default cost assessment process. For base expenditure, this will relate to the coverage of the identified expenditure within the final suite of base econometric models used. We believe that it is a critical component of setting out the case for, and assessing, cost adjustment claims. They require a clear awareness of the scope and performance of the econometric models that will be used. Therefore, the potential early publication of models will help to facilitate better understanding and encourage companies to focus efforts on evidencing cost adjustment claims that remain material after an implicit allowance has been considered (a net cost adjustment claim).

We accept that the calculation of implicit allowances can be complex; but found that it was sometimes difficult to follow and therefore learn from the process that had been undertaken to calculate implicit allowances at PR19. Therefore, we would encourage Ofwat to set out in good time ahead of PR24 cost adjustment claim submissions, some principles and examples of how IA's should be calculated in typical cost adjustment areas, and how this interacts with the 'need for cost adjustment' assessment gate. This will allow companies to refine and prioritise potential claims accordingly. Similarly, it appears sensible to encourage companies to set out their cost adjustment claims both net and gross of IA including the assumptions used in arriving at a material net cost adjustment claim.

Timing of cost adjustment claim submissions

We cautiously welcome the opportunity to submit cost adjustment claims in advance of the final business plan submission. This is particularly relevant if symmetrical claims are likely to be used more broadly. In this case, early visibility of models will only be of major benefit if there is also early visibility of likely symmetrical claims (as together they are going to inform the likely cost benchmark that company plans will be assessed against).

The timing of an early submission is important. Feedback on and visibility of the type of claims being proposed would be important. But this must be on a timescale whereby there is an appropriate opportunity to respond to and incorporate any feedback. An equivalent timescale as at PR19 (May 2023) feels too close to the final submission deadline for substantive reassessments to be undertaken. Similarly, submission of claims too early will likely reduce the quality of the claims and increase the disconnect with the final submitted plan.

Tight P consents – the potential case for a non-symmetrical adjustment

In ***Chapter 5*** we discuss the need to better reflect future capital maintenance and asset health pressures driven by externalities and enhancements. An example of this is the future operation and maintenance of sewage treatment works with very tight Phosphorous consents. Where modelling cannot reflect these pressures (as discussed below), non-symmetrical adjustments would be an appropriate mitigation.

Prior to AMP7, the majority of Phosphorous (P) removal consents did not require phosphorus below 0.5mg/l. This will change throughout AMP7, as shown in ***Figure 4.2***. We expect to see dramatically increasing levels of tighter consents for phosphorus removal and a radical reduction in the amount of waste discharges without phosphorus consent.

This needs to be accounted for but there is relatively little historical data for P-Consents below 0.5mg/l. This means we cannot feasibly include it in the econometric specifications despite clear engineering logic that we expect to see higher costs associated with this tightening of consents.

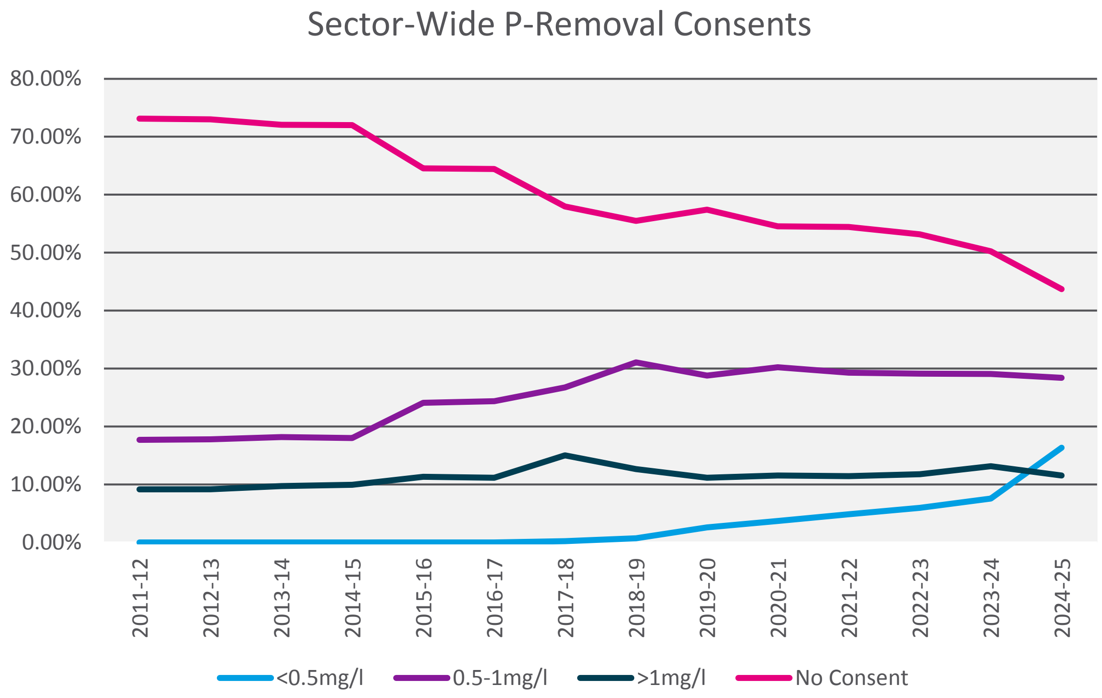
Figure 4.1 shows an amended SWT1 model with 'pctnh3below3mg' being replaced with 'pctpbelow05mg', a variable describing the percentage of load treated with a phosphorous consent below 0.5mg/l. We can see that the RESET test is just satisfied, and the regression even gives a higher R^2, and better AIC and BIC values. It also satisfies engineering logic – we can feasibly suggest that tightening P-consents will be associated with increased costs.

Figure 4.1 – Regression coefficients (p-values) for the original SWC1 specification and a new specification replacing pctnh3below3mg with pctpbelow05mg. These use a panel of data from 2011-12 to 2019-20.

Variable	Original SWT1 Model	SWT1 with pctpbelow05mg
lnload	0.7794 (0.000)	0.8888 (0.000)
pctbands13	0.0416 (0.016)	0.0427 (0.008)
pctnh3below3mg	0.0036 (0.000)	-
pctpbelow05mg	-	0.0358 (0.227)
constant	-5.2144 (0.000)	-6.5302 (0.000)
R^2	0.8728	0.8844
AIC	-100.81	-103.41
BIC	-85.81	-88.41
RESET	0.3106	0.0502

This highlights a need to be careful. All the statistics would suggest that this is a perfectly reasonable regression to use. When business plan forecasts are used as the forecast drivers, however, this would raise the sector's wholesale base allowance by more than £1.5bn. The increased costs from tightened P-consents are unlikely to amount to such a significant sum, but there will be associated costs. This isn't something that can be easily modelled, making this a candidate for a non-symmetrical cost adjustment applied to the whole sector.

Figure 4.2 – Changing sector-wide P-Consents between 2011-12 and 2024-25.



Potential cost adjustment claim subjects

We set out below a list of potential adjustments that we might consider developing and putting forward at PR24. These largely reflect the issues that we have discussed in ***Chapter 3.3*** relating to model specification, and the future cost pressures associated with ongoing maintenance as discussed in ***Chapter 5*** (e.g. operating low P-consent sewage works in the future that is discussed above).

Separately, we are also considering the potential scope for enhancement business cases and associated enhancement cost adjustment claims. We are aware that enhancement cost assessment is not within the scope of this consultation and are engaging separately with Ofwat in this area through the CAWG.

Where model specifications change to adequately address these identified issues, the adjustments would not be required. Equally, as we develop our business plan and gain greater insight of the cost assessment / modelling approach that will be used at PR24, other potential cost adjustment claims may arise. Therefore, we do not consider that this current list is necessarily a long list which can only be refined down.

For the model specification issues that we have identified, we have commissioned consultancy support that will consider in more detail the empirical evidence to support the higher-level engineering logic that we have raised. This will help to improve our understanding of how cost drivers directly influence costs. This will support both the evidence to justify the need for, and potentially a way of implementing/scaling, any adjustment.

Potential claim subjects	High level basis	Symmetrical / non symmetrical	Possible data requirements
Density / sparsity	Areas of high density and sparsity both drive cost, but this is not captured in the weighted average density variable currently considered in the base models.	Symmetrical	Population density information at a more granular level (e.g. postcode sector data from census)
Water treatment complexity	Water treatment costs are driven by whether the source is ground or surface water and the processes that are installed at those sites. Our analysis suggests that the current water treatment complexity data may not adequately reflect these drivers of cost.	Symmetrical	Operating and capital maintenance information by water treatment works (or assemblages of works). Analogous to Large STW table
Water network complexity	Depending on the network complexity explanatory factors used, the operating or capital maintenance costs are not likely to be appropriately accounted for. These are material costs which are not necessarily sensitive to one cost driver	Symmetrical	Standardised APH data Booster pumping station data
Water Resources costs	The type, location and size of water resources assets impact on the costs that companies incur in both the water resources and network plus price control (through the knock-on impact on treatment complexity). These are controlled largely by the geography and geology of the company rather than scale or population density.	Symmetrical	More granular water resources related cost driver information. For example, expanding the information previously reported at PR19 in Wr2 to account for capex.
Future operating and maintenance impacts of prior enhancement expenditure	Historical costs will not account for the future operation or maintenance of enhancement activity implemented in the previous period (e.g. WINEP driven activity). The future impacts of enhancement interventions are unlikely to be strongly sensitive to model scale drivers. Consequently, over and above any allowance through a time trend or sensitivity to a complexity driver, very little allowance can be assumed.	Non-symmetrical	Whole-life cost information relating to incurred enhancement costs expenditure
Operating Sewage Treatment Works with Tight P consents	Operating and maintaining Sewage Treatment Works with tight P consents will have different cost characteristics than seen historically. There are also likely to be material knock-on impacts to bioresources costs. Therefore, backwards looking models are not likely to adequately reflect these costs.	Non-symmetrical	Relative costs of operating and maintaining STWs across the full spectrum of P consent levels
Biodiversity net gain	The Environmental Act (which has recently received Royal assent) sets out a future requirement to deliver a 10% biodiversity net gain when planning permission is required. This will create a new cost pressure when delivering major capital maintenance projects.	Non-symmetrical	Biodiversity activities delivered associated with the capital programme

5. Capital Maintenance and Asset Health

Summary of content / messaging in Ofwat's consultation document

- Ofwat does not believe there is yet a case to change its approach to funding as a result of future asset health cost pressures. It suggests that it has not yet seen sufficient evidence from companies on how future capital maintenance requirements will change.
- However, Ofwat accepts the CMA's suggestion that it should consider forward looking indicators in the cost assessment process to aid in assessing future capital maintenance expenditure.
- Ofwat provides a number of ways to implement this forward-look into the process. But at present it does not consider that asset health metrics should be added as explanatory variables in the wholesale base cost models as this would risk introducing endogeneity into the models.
- Ofwat is interested in the creation of a wider range of asset health measures to form a more holistic view of the state of asset health across the sector. This will be used to ensure cost allocations are adequate, to hold companies to account with respect to the health of their assets and provide incentives for long term resilience.

Consultation Questions asked
19. Do you agree with the different elements / approaches to introducing more of a 'forward-look' into our approach to assessing capital maintenance expenditure? Are there other elements / approaches we could consider?
20. Do you have any comments on the proposed long list of asset health measures, particularly in relation to suitability and how feasible they are to collect? Please include any reporting or definition changes you would like us to consider and provide suggestions for other measures not included in this list.

Severn Trent Response

In its consultation, Ofwat states that "the evidence available to us does not suggest that [future asset health pressure] is an issue that requires us to amend our cost assessment approach."

We think Ofwat could take some more time to consider whether there is a material asset health challenge in the future that will need addressing. Emerging future pressures will impact on the size and shape of the asset base we rely on, the amount of stress that it will likely be subject to, and the way in which it will respond to those stresses. These include:

- increasing climate change uncertainty and extremes;
- customer driven enhancements (e.g. resilience and wider environmental enhancements (such as river pollution));
- newly introduced legislation such as biodiversity net gain;
- long term water resources and drainage management planning objectives;
- delivering against net zero objectives; and
- responding to new and emerging risks such as cyber security threats.

They will all drive future operating and maintenance cost pressures.

We acknowledge the complexity of the interaction between asset health and capital maintenance investment. It is overly simplistic to consider that investment in asset replacement has a one for one relationship with asset health. This is illustrated in the comparative information that has been presented in the consultation document:

- Asset health measures (showing improvement over the historical record)
- Asset replacement rates (showing reductions over the short/medium term)
- Capital maintenance expenditure information (showing relative increases)

Major complicating factors can be attributed to the changing contextual environment in which our assets operate (assets operating under greater stress will have different patterns of deterioration); and more sophisticated methods of operating and improving assets or more effectively responding to asset failure before it affects service, which companies have increasingly deployed.

More sophisticated asset management methods include network calming, pressure management and more intelligent maintenance monitoring and scheduling. All these activities will likely extend the life of traditional assets and manage the impacts of failure on service to customers. However, they do not negate the need for asset replacement/renewal in the longer term.

Therefore, we believe that a key objective for a cost assessment approach should be to make sure that adequate provision for capital maintenance and network renewal is made that accounts for:

- the physical properties and changes in the asset base over time;
- the external pressures that may be increasing the amount of stress on the asset base in the future;
- improvements that have been made in managing existing assets more effectively; and
- the need to counter potential natural incentives to chase shorter term cost efficiencies that may exist if cost assessment and asset health deliverables are not closely aligned.

Considering the above observations, and Ofwat's concern on the lack of supporting evidence, we support the desire to improve understanding on asset health. The materiality of associated investment and its fundamental role in delivering future core service mean that it is particularly important that regulatory decisions in this area are based on robust and well understood comparative data.

However, it is not yet very clear how newly collected asset health information will materially inform the PR24 base cost assessment. Given the likely increases in regulatory burden, the specific types of evidence that Ofwat would consider useful to improve understanding and confidence will need to be carefully considered.

5.1 Potential approaches to improve consideration of future challenges

The consultation document sets out three systematic options for how future capital maintenance requirements might be better accounted for in the cost assessment process.

- Using business plan forecasts as inputs into base econometric models – As already stated in ***Chapter 3.4***, **we do not support this option**. We consider the use of business plan forecasts would materially reduce the robustness of models and therefore the accuracy of any efficiency benchmark.
- Setting the catchup efficiency challenge based on business plan forecasts rather than historical outturn costs – **We cautiously support** more consideration of this option, but think it would need to be carefully managed for it to deliver as intended.

- Including explanatory factors to the models that directly relate to capital maintenance / asset health activity – Whilst we are aware of Ofwat's concerns about endogeneity, **we think that this option still has merit** particularly if combined with independent assessments of the endogenous drivers and clear linkage to the outcomes framework to ensure delivery of long term asset health / capital maintenance deliverables.

We discuss these options in more detail below.

We do not consider that separate modelling of capital maintenance is a realistic option as it is counter to totex assessment which underpins the whole cost assessment approach.

In addition to these systematic options, future challenges can also be accounted for through the use of non-symmetrical cost adjustment claims as discussed in ***chapter 4.***

Including business plan forecasts in cost models

This option has some theoretical attraction. Assuming that maintaining future asset health will become increasingly more challenging, this should manifest in increased base expenditure being seen in business plan submissions. Inclusion of these forecasts in models would then mean that these increased asset health pressures would start to be accounted for in the model outputs.

However, we consider that such an approach would lead to a significant increase in uncertainty. Therefore, we do not think it is viable to include forecast base costs in base cost models.

- The additional levels of uncertainty in cost forecasts relative to outturn costs will significantly reduce the robustness of the models. We have set out analysis in ***Chapter 3.4*** that illustrates the detrimental impacts on model robustness of including PR19 forecasts in current models.
- In addition to the inherent levels of cost uncertainty associated with using forecasts, additional uncertainty is likely to transpire from the likelihood that companies will have different price review strategies / attitudes to risk; use of ODIs; and exposures to future cost pressures. These would also contribute to increased model noise.
- Future external pressures are likely to become increasingly apparent across the extended panel. Therefore, without the inclusion of a time trend explanatory variable, the model coefficients would seek to average out these impacts, thus creating a misspecification. We also note that assuming this trend is already in evidence, the use of a time trend explanatory factor could take account of the future pressures without the complicating effects of using forecasts. We have discussed the case for acknowledging time effects in the models in detail in ***Chapter 3.3***.

If forecast costs were to be included in models, appropriate acknowledgement of the increased uncertainty inherent in the model would be necessary through a reduction in the strength of the catchup efficiency challenge applied.

Setting the efficiency challenge on the business plan forecasts

Traditionally, the setting of the catchup efficiency challenge relates to the level of confidence in the predictive capabilities of the model. The selection of the benchmark company/companies is currently an exercise in removing the effect of outlier companies whose calculated efficiency performance may be overly impacted by model uncertainty.

This option would increase the function of the catchup efficiency challenge to also address future capital maintenance pressures not appropriately captured in the historic dataset. The key benefit of this option relative to using forecasts in models is that the coefficients will be based on the observed relationship between cost and cost driver without the increased uncertainty of using forecast data.

Assuming company forecasts are closely related to their observed efficiency performance, and have commensurately accounted for future asset health pressures, the distribution of company efficiency scores using this approach should stay equivalent to an historical calculation. The magnitude of the challenge would then change to account for the future pressures faced. Depending on the extent of the future pressures, this could either reduce the strength of the catchup efficiency challenge or turn it negative (i.e. increasing rather than reducing the size of pre-efficiency model forecasts).

However, both assumptions (linkage of outturn efficiency to plan, and plan to future health pressures) will introduce uncertainty to the efficiency catchup calculation. Consequently, careful consideration of the selection of the benchmark company/companies would be required. This might be through a reduced challenge to account for the increased uncertainty (e.g. moving away from UQ towards the median), or having a more bespoke consideration of the applicability of companies to be selected as the benchmark.

Bespoke consideration would need to consider whether submitted plans may be overly aggressive or short-termist. This could be informed by reviewing the delivery of historical and forecast asset health commitments and reviewing the variance of equivalent efficiency scores derived from outturn data and forecast estimates. Such rigorous checks and balances would help to give confidence that selected benchmarks are not overly aggressive / short-termist and likely to promote a sustainable level of asset health in the future.

Therefore, this option is effectively:

- distributing historical costs between companies based on the explanatory factors derived from observed data; then
- scaling these outputs so that they correspond with business plan submissions which can be demonstrated to be robustly delivering future asset health through a sustainable capital maintenance plan; and
- appropriately accounting for uncertainty within the efficiency score calculation.

The implementation of this option could be undertaken at a sector level (replacing the existing calculation of catchup efficiency using outturn costs), or by way of a more targeted trial whereby companies become eligible for a forecast catch-up efficiency adjustment upon demonstrating a level historical asset health competency and delivery.

Including more explanatory factors in models e.g. renewal rate or network age.

We note that Ofwat have stated that they are not likely to not pursue this option using common asset replacement metrics. This is due to concerns with endogeneity and potential for double funding. We can understand the reasoning behind wanting to avoid including endogenous explanatory factors in cost models. Cost models are attempting to understand the distribution of costs once the external factors that are driving costs have been taken account into account. From this perspective, it does not make sense to include measures of asset health as model explanatory factors.

Replacement rates and asset lives are clearly subject to company choice and risk appetite. However, they are also impacted by the cost efficiency incentives of the price review which drive companies to reduce cost subject to them also being able to deliver against their stated performance commitments. Where performance can be maintained in the shorter term through operating / remedial solutions but potentially to the detriment of longer-term asset health, the current regulatory framework does not discriminate.

Including appropriate explanatory factors such as renewal rate or network age would provide the clearest way of counteracting the above adverse effect of the cost sharing incentive, ensuring that long term future asset health pressures are being adequately incentivised. We also think that some of the perceived issues can be managed through independently setting forecasts for the explanatory factors and strengthening the linkages to performance commitments. Therefore, we would encourage Ofwat to consider how this approach might be delivered with appropriate mitigations also in place.

To implement this option, historical asset heath metrics (e.g. including network renewal or asset failure rates) would be included when creating models. This would expose the cost relationship associated with delivering investment in a way that will promote long term sustainable asset health. When using the model to forecast the future expenditure requirement, the asset health explanatory factor forecasts could then be stipulated at a required level. This might be set according to:

- engineering expectations (e.g. a maximum plausible operating age);
- comparatively (industry average or upper quartile); or
- at an uncontentious minimum level (then linked to an ODI that would allow further investment where higher levels of activity are economic).

To protect the interests of customers and ensure future pressures are appropriately managed, companies would then be required to deliver in line with the model inputs by also making the explanatory driver forecasts a performance commitment.

The consultation document also questions whether it might be more appropriate to include asset base drivers where circumstances have changed, or new data has become available. We would caution against using forecast drivers that have investment that is either significantly smaller or not included in the historical data set. In these cases, the relationship between cost and cost driver that will be calculated during model construction is likely to carry much less predictive capability for the future anticipated volumes. We have demonstrated this with respect to using a tighter P-consent level in the wastewater models in ***Chapter 5***. In these cases, we consider that a non-symmetrical modelling adjustment would be a much more appropriate way of accounting for these material future costs.

5.2 Asset Health Measures

Identifying and developing potential new asset health metrics

The criteria for assessing potential asset health metrics that is set out in table 5.1 of the consultation document is sensible and coherent. We explicitly support the acknowledgement that back-casting new asset health metrics may be challenging where they have not been specifically collected in the past. We also think that Ofwat should prioritise data that has the greatest likelihood of providing robust comparative information and be mindful of the regulatory burden of using complex metrics that require significant analysis to collate and interpret.

We do not think it is sensible to pre-empt the UKWIR work on potential asset health measures. However, the four proposed types of asset health measures[2] appear sound. As to the collated examples of potential measure across each of the four types that are presented in table 5.2 of the consultation document, we make the following high-level comments:

[2] Asset Characteristics, Maintenance activity, Asset and service performance measures and Aggregated measures.

- Care should be taken when using service performance measures to make inferences on underlying asset health – This is because it can be possible to maintain service/ performance in the short term through shorter term operating type solutions which may or may not have an impact on the underlying long-term performance and capabilities of the asset base.
- Measures of maintenance activity are likely to be subject to management control and would therefore need to be treated with appropriate caution.
- It may be sensible to consider the use of metrics that better track changes in the shape and characteristics of the asset base (e.g. number / value of short life EM&I (Electrical, Mechanical and Instrumentation) assets including meters, loggers etc). These could potentially help to indicate changes in future maintenance requirements.
- Some of the aggregated measures can be very complex to calculate and require significant analytical work to report. We have developed the Overall Equipment Effectiveness (OEE) measure. This uses a wide range of measures at a granular level to reveal patterns of asset performance by different asset components or locations. OEE analysis is relatively new to us but is likely to be a very powerful tool for understanding and prioritising our asset management work. However, this could be very challenging to convert into a robust comparative measure between companies (it would likely require many thousands of observations). It would also lose much of its value if considered only at a very high level. Therefore, we suggest that this might not be a very transparent cross sector measure.
- There may be merit in focusing asset health measures where there are known variances with respect to company asset bases (e.g. network assets in rural areas) or where asset renewal is expensive or lumpy (e.g. trunk main renewal or DSR (Distribution Service Reservoir) replacement)

It is not yet very clear how newly collected asset health information will materially inform PR24 base cost assessment. Where the data is not already collected to an appropriate level of maturity, we do not think that inputs into models are very likely. It is also unclear how Ofwat propose to use the information outside of models. The consultation document states both that this information would not form PR24 performance commitments, but also that it would be used to hold companies to account for comparative asset health performance.

Most of the identified metrics in table 5.2 have a level of prior reporting. This will be important if consistent comparative information is to be generated. Where this is the case, prior reporting definitions should form the starting point for developing clear reporting requirements for these measures. However, for information that has not previously been reported (i.e. used for internal purposes only) a series of iterations to refine guidance, and company interpretation of that guidance, will be necessary before meaningful comparisons can be expected.

It will also be important to consider how these metrics are likely to perform in the future given anticipated increases in environmental pressures. It would be too simplistic to project metrics forward assuming that assets will be operating in a steady state context (e.g. the impacts of climate change, population, extreme events, increasing expectations, changing asset base characteristics).

6. Cost-service relationship

Summary of content / messaging in Ofwat's consultation document

- Ofwat is considering forming a more explicit link between cost and service quality at PR24, and using this to aid in the setting of cost allowances and performance commitments.
- The key question it asks here is 'What does base buy?'.
- It is interested in evaluating whether performance commitments should be common or company specific, determining the appropriate baseline performance for an efficient company in 2024-25 and forecasting this performance through to 2029-30.
- It assumes that:
  - Performance commitments should be common, unless there is robust evidence they should be set on a company-specific basis
  - That efficient companies will on average deliver their performance commitments
  - That efficient companies will continue to improve performance in the long term
  - That it will need to review the performance level that base buys at each price review to account for changes such as technological change.
- It considers answering the 'what does base buy' question by looking at historical outturn data, PR19 performance commitments, PR19 business plans, and PR24 business plans. It will consider using this data to create a performance trend.
- It also discusses the viability of econometric approached to explore the cost-service relationship.

Consultation Questions asked

20. Do you agree with the high-level approach to determine 'what base buys?' Can you define any additional analysis or information that could support this process?
21. Do you consider it would be feasible to assess the 'efficient' baseline performance level for each company for individual PCs such as leakage and PCC through econometric modelling? Are there any other PCs where you consider this could feasibly be attempted?
22. The need to collect further granular data to elucidate the cost-service relationship was highlighted by companies in response to our PR24 May consultation. Can you propose any data it would be appropriate to collect to support the high-level approach outlined in the chapter?
23. What are your views on attempting to use a composite variable to investigate the cost-service relationship, in the context of the methodological issues and complexities we outlined?
24. Do you have any proposals for how to make adjustments where a performance commitment level differs from that expected to be delivered from base costs?

Severn Trent Response

We agree that the interaction between cost and performance is a fundamental component of the price review. At PR19, econometric cost models essentially assumed that companies were delivering a common level of performance. Where this was not the case, and these performance variations resulted in different levels of cost being incurred, the discrepancy will have contributed to model noise. This uncertainty should be accounted for when setting the strength of the catch-up efficiency challenge.

However, we think that the logic set out in section 6 of the consultation document requires further consideration and may lead to some undesirable outcomes. We are concerned that the fundamental definitions and expectations of base and enhancement expenditure may be overridden if some of the proposals set out in the consultation are implemented. This would lead to a confused interaction between base and enhancement activity and unnecessary complexity.

Therefore, we would urge Ofwat to make sure that the costs that are being included within the scope of base econometric models reflect - and are accompanied by - a realistic assessment of the outcomes that they funded (and appropriate recognition of how the operating circumstances within which outcomes have been delivered in the past might be changing, most notably, because of climate change pressures). We think that Ofwat should apply appropriate remedies where it considers the interaction between cost and performance risks is giving rise to an unacceptably large level of model misspecification.

Base costs should relate to maintaining a constant level of base service

We do not agree with the statement that "efficient companies will continue to improve performance over the long term from base expenditure" given the definition of base costs.

Base costs relate to capital maintenance and base operating expenditure. This is defined as the expenditure needed to maintain a constant level of performance. It is stated in the reporting requirements set out in RAG4.10 (Table 4J and K):

- "Maintaining the long-term capability of the assets: Capital expenditure on [infrastructure / non-infrastructure assets] excluding third party capex to maintain the long-term capability of assets and to deliver base levels of service. Where projects have drivers both of enhancement and capital maintenance, companies should apply a method of proportional allocation to allocate costs between enhancement and capital maintenance"

We consider that this distinction has been the case across the duration of the data panel. Prior to AMP6, capital maintenance was defined as the level of expenditure to deliver the reference level of service (as set out through the previous serviceability methodology)[3].

We agree that service levels might improve through time as companies identify new ways of working and respond to ODI incentives. Where this is the case, Ofwat would rightly want to protect the interests of customers by holding companies to those improvements at successive price reviews where they are demonstrably beyond interannual variation. This is like the ratchetting mechanism of the previous serviceability reference level methodology.

However, whilst companies can choose to improve levels of service, this should not be considered as a zero-cost efficiency improvement (i.e. in most cases, companies would need to choose to forego cost savings to drive performance improvements). This would typically be facilitated through the interaction of cost allowances and bespoke ODI incentives[4] and will create an ongoing maintenance pressure to maintain these new service levels. Consequently, we are strongly of the view that a rolling level of performance improvement driven from base expenditure should be considered as a corporate choice rather than a starting assumption.

---

[3] As a default the reference level of service was considered to be the average of the best level of performance previously delivered and its nearest neighbouring year. Reference levels were only set greater than the historical performance level where corresponding enhancement expenditure was also being incurred.

[4] As we mention in our other consultation responses, we consider that bespoke ODIs should remain a vital part of the outcomes framework because they allow companies to innovate with different service improvements.

We can also accept the basis that companies who do not meet a defined level of service should be expected to improve. For example, a common service level where there are no external drivers such as regional geography or geology that would suggest the different levels should be expected. This is effectively saying that low performing companies are catching up to better companies. We consider that this is a different logic to assuming that base expenditure will on its own lead to a gradual increase in service performance.

The cost-service assessment framework – 'What base buys?' and subsequent use of adjustments

We support the aspiration to identify the level of performance associated with modelled base expenditure (i.e. what does base buy?). However, we do not believe that the approach proposed is appropriate.

The base econometric models account for all relevant base operating and capital maintenance expenditure. This means that the models take account of all performance levels across the sector (i.e. the full spectrum from best performing company to worst performing company). The models seek to identify the relationship between costs and fundamental cost drivers. At PR19, performance metrics were not included as model drivers due to concerns with endogeneity[5]. A fundamental consequence of this is that the coefficients will assume an average level of performance / service provision (i.e. because the costs of all levels of performance have been included in the data panel).

Assuming that there is a recognisable relationship between cost and performance, when the catchup efficiency challenge is set, the corresponding performance assumed should therefore relate to the performance of the company/companies that are setting this efficiency challenge (e.g. 4th company, or the company/companies that are at UQ). This is not the same as defining the upper quartile performance level which may or may not be driven by differential levels of expenditure.

The consultation document asserts that a relationship between historical cost efficiency and outcomes performance can be identified. However, our own analysis suggests that this relationship is poorly defined and circumstantial – while there are examples of cost-efficient companies exhibiting high levels of performance, this is not seen consistently. This is also supported by the CMA findings. Therefore, when identifying the level of base performance associated with the efficiency benchmark, it would be essential to consider a range of company performances (e.g. +/- 1 or 2 companies around the benchmark).

When reviewed against PR19 data, our analysis shows that performance varies significantly with cost efficiency. With an appropriate level of tolerance, the performance of the benchmark companies' quickly tends towards the average. This supports the hypothesis that performance is also subject to external factors not accounted for in model explanatory factors, rather than having a one-to-one relationship between efficiency and performance.

In conclusion, the starting point for defining the level of performance assumed in the base econometric models (i.e. What does base buy?) should be the industry average level of performance. Continuing this logic, the identified average level of performance should be projected forward on a

[5] It is possible but not proven the some of the chosen exogenous cost drivers (such as density or complexity) may indirectly contribute to performance levels. The engineering basis for this has not been explored in detail. In principle, we consider that the decision as to whether performance should be focused to a common or bespoke service level should be based on the extent to which performance can be shown to be driven by exogenous factors outside of company control (bespoke) or not (common).

flat basis (or sensitive to any related enhancement expenditure assumed at PR24) rather than following an ongoing trend.

To protect the interests of customers, we are aware that Ofwat should scrutinise where performance materially diverges from the average. This could include the removal of underperforming companies from consideration when calculating the efficiency benchmark. Where this relates to model specification, this could be through specific/symmetrical adjustments or ODIs. However, we do not feel that diluting the fundamental principles of base and enhancement expenditure and adding unnecessary complexity to the efficiency methodology is a transparent or coherent approach to follow.

Using econometric models to identify levels of performance

Econometric analysis could theoretically be used to identify efficient levels of identified base service. Rather than cost, the dependent variable would be flipped to a particular aspect of performance. This would then be explained through external explanatory factors that will impact on performance outside of company control. Given the rationale of common performance levels (i.e. identified targets are applicable to all companies), this econometric approach would only be justifiable for non-common performance levels.

However, like cost modelling, such an approach would be subject to major uncertainty. We do not believe that it is realistic that complex performance metrics (such as leakage or PCC) can be adequately developed to an appropriate level of robustness at PR24.

Models would require an equivalent level of engineering logic as required for cost models. It is not clear whether performance could be simplified down to several explanatory variables using data that has not yet been defined and refined for consistency. When coupled to the existing uncertainties originating from cost modelling, it is unlikely that an appropriate level of robustness could be provided for. The complexity of such a composite modelling approach could also significantly reduce the level of transparency and does not feel consistent with the aspiration for sensibly simple models. This is particularly true for performance measures like leakage where local factors might play an important part in performance, and it might be difficult to create company-wide variables to explain the variation.

We note the potential use of composite performance variables as a way of including performance cost drivers into cost models. This may help to improve model specification, and principal component analysis could be used to avoid the need to define particular weightings (as per ***Chapter 3.3***). Endogeneity issues could also be managed by using independently defined forecasts (as per ***Chapter 5***). However, the aggregation of metrics would only improve model predictive performance if the performance metrics were broadly consistent, and sensitive to the same underlying exogenous explanatory factors. Therefore, before such approaches were pursued, we suggest that the causal relationships between cost and cost drivers for each of the of performance variables would need to be mapped out and explored in more detail.

7. Residential retail cost assessment

Summary of content / messaging in Ofwat's consultation document

- Ofwat welcomes feedback on the cost drivers used for retail cost assessment at PR19 and are open to suggestions for any new variables to be considered at PR24.
- It intends to ask companies to separate out any Covid-19 related bad debt costs from their standard provisions.

Consultation Questions asked

20. Do you have any comments regarding our proposal to ask companies to separate out the part of their provision of bad debt costs to do with Covid-19 that was made outside of their standard methodology in the PR24 business plan tables?
21. What guidance would aid companies to provide appropriate data related to the provision of bad debt costs to do with Covid-19?

Severn Trent Response

We are supportive of the intention to separate Covid-19 related bad debt costs from the standard provision, but question whether this might lead to issues with the data. Relevant guidance would be sensible to ensure consistency in reporting across companies.

If it is not possible to effectively remove Covid-19 bad debt costs the exceptional year indicators discussed in ***Chapter 3.3*** may be a better solution to this problem than any smoothing or making no adjustment. These indicators will account for non-repeatable costs in exceptional years – i.e. an indicator equal to 1 for the pandemic years and 0 otherwise.

Smoothing will only serve to spread the 'noise' across multiple years, while making no adjustment will keep the noise in the model for the 'covid years'. Using a time indicator should help to eliminate the majority of the noise while keeping years not impacted by Covid-19 consistent.

We would also welcome the inclusion of a broader range of affordability measures, which will be available from the credit ratings agencies – e.g. energy bill sizes, housing costs – given the cost of living increases that customers are currently facing.

We look forward to engaging with Ofwat on residential retail cost assessment in more detail ahead of the draft methodology in July. We consider there is no clear rationale for residential retail being a non-indexed control. We note that in the Retail Exit Code protections for small non-households (those most like residential customers), Ofwat decided to apply an indexing control, which makes the continuation of the current non-indexed approach for residential retail inconsistent with the market. For non-household retailers, Ofwat acknowledged that inflationary pressures applied and that the simplest way to deal with these was by indexing costs in line with CPIH.

Appendix A: Supporting evidence on wholesale base cost drivers and explanatory variables

A.1 Potential Wastewater Network Plus model specification

We have started to consider potential model specifications for wastewater network plus models. To do this, we have considered the engineering logic and statistical performance principles set out in ***Chapter 2.1*** and ***Chapter 2.2*** above.

It has been difficult to come up with a set of variables that account for the principal cost drivers that we should expect, are statistically significant, satisfy the RESET test for specification and the Variance Inflation Factor test for collinearity.

We summarise the thinking behind our model development to date with reference to the following principal cost drivers.

- **Scale** – The need for a robust scale driver is self-explanatory. However, this is complicated by the fact that the logical scale variable for the SWC component is sewer length, and for the SWT component is load. Therefore, we have taken this opportunity to explore the use of Principal Component Analysis (PCA) to make a composite scale driver that takes account of both sewer length and load. The basis for composite scale drivers and PCA is described in ***Chapter 3.3*** and in more detail in ***Appendix A.7***.
- **Density** – Areas with high population density can lead to congestion effects in sewage collection but can also enable economies of scale in treatment. The economies of scale can be accounted for directly, so we would expect measures of density to have a positive coefficient so long as the model is properly specified. In ***Chapter 3.3*** we undertake some detailed analysis of density metrics and their impact on costs. These arguments – centred on potential concerns about the use of weighted average local authority data – are relevant here too. Therefore, we could alternatively specify an explanatory variable for density as follows:
    - a percentage of area above a certain density threshold,
    - a weighted average,
    - as properties per length of sewer etc.
- **Network Complexity** – This considers measures that proxy for topography. Possible options are pumping capacity per sewer length or rising mains as a percentage of total sewers. It would be attempting to account for increased sewage collection costs driven by pumping needed to overcome gravity. In all specifications attempted, these variables were insignificant and reduced the quality of the models and therefore been removed. Further consideration of the data may be sensible. However, given that all the costs should sit within the network it stands to reason that network complexity will be less material in Wastewater Network plus model scopes than Sewage Collection. Additionally, pumping provision is much smaller in wastewater than in water networks meaning that the need for appropriate network complexity drivers in Wholesale Water and TWD models is stronger.
- **Treatment Complexity** – These are variables that account for increased sewage treatment costs because of stricter consents for wastewater coming out of treatment works. We have data on BOD, P and NH3, and can use any of these or a composite of them.

- **Economies of Scale** – these are measures of anything that might be able to give benefit from economies of scale. Most obviously is the ability for companies to have large sewage treatment works. The 'pctbands13' and 'pctbands6' variables used at PR19 account for this.

Figure A.1.1 – Regression coefficients (p-values) from various proposed NPWW specifications, using a panel of data from 2011-12 to 2019-20.

Variable	NPWW1	NPWW2	NPWW3
pca_scale	0.3228 (0.000)	-	-
lnload	-	0.7182 (0.000)	-
lnsewerlength	-	-	0.7549 (0.000)
lndensity	1.1984 (0.083)	0.8721 (0.178)	1.4777 (0.035)
pctnh3below3mg	0.0030 (0.000)	0.0029 (0.001)	0.0032 (0.000)
pctbands6	-0.0154 (0.017)	-0.0144 (0.027)	-0.0138 (0.039)
constant	2.2051 (0.317)	-5.7606 (0.000)	-7.1324 (0.005)
R^2	0.9095	0.9145	0.9069
AIC	-160.34	-158.15	-160.30
BIC	-142.84	-140.65	-142.80
RESET	0.3021	0.4362	0.2291

Some caution is needed though given potential collinearities that we have identified across the current drivers. This is particularly relevant here given potentially conflicting cost drivers between treatment and collection.

Therefore, further work could be to consider alternative variables (particularly those not standardised by length).

Acknowledging the potential increasing challenges of interpretability and validation, a greater use of PCA may be an appropriate alternative to consider. The PCA could be calculated to account for all variables (i.e. increasing the PCA scope to account for scale, density and complexity) that are supported by engineering logic. Then additional principal components of the PCA could be included into the model specification rather than relying on separate cost drivers. A regression showing this approach is displayed in ***Figure A.1.2***, and appears to better fit the data than the regressions in ***Figure A.1.1*** by AIC, BIC and R^2 while accounting for more potential cost drivers.

Variable	NPWW PCA
pca1	0.1773 (0.000)
pca2	-0.0339 (0.388)
pca3	0.2445 (0.000)
pca4	0.1996 (0.002)
constant	5.5622 (0.000)
R^2	0.9388
AIC	-163.51
BIC	-146.01
RESET	0.0542

Figure A.1.2 – Regression coefficients (p-values) from various an NPWW specification, using the first four principal components from an analysis over lnload, lsewerlength, lnpumpingcapperlength, lndensity, lnwedensitywastewater, pctnh3below3mg, pctbands6 and pctbands13. This uses a panel of data from 2011-12 to 2019-20.

A.2 Water Network Complexity

Figure A.2.1 – Regression coefficients (p-values) with just BPL, APH + BPL, just APH for the WW1 specification. These use a panel of data from 2011-12 to 2019-20.

Variable	Original WW1 Model	WW1 with APH + BPL	WW1 with just APH
lnproperties	1.0365 (0.000)	1.0442 (0.000)	1.0578 (0.000)
pctwatertreated36	0.0055 (0.000)	0.0048 (0.001)	0.0046 (0.006)
lnwedensitywater	-2.3239 (0.000)	-2.2408 (0.000)	-2.6690 (0.000)
lnwedensitywater2	0.1652 (0.000)	0.1598 (0.000)	0.1839 (0.000)
lnboostersperlength	0.3531 (0.005)	0.3313 (0.009)	-
lnaph_ww	-	0.1475 (0.176)	0.1676 (0.099)
constant	-0.3734 (0.777)	-1.5509 (0.326)	-1.4047 (0.501)
R^2	0.9732	0.9749	0.9716
AIC	-139.11	-140.11	-134.35
BIC	-114.61	-112.55	-109.85
RESET	0.1276	0.0960	0.2261
MAPE 2018-20	14.98%	14.05%	14.96%

Figure A.2.2 – Regression coefficients (p-values) with just BPL, APH + BPL, just APH for the WW2 specification. These use a panel of data from 2011-12 to 2019-20. Wholesale APH is calculated as a sum of APH in WR, RWD, WT, and TWD. This assumes an equal volume of water entering each control is identical.

Variable	Original WW2 Model	WW2 with APH + BPL	WW2 with just APH
lnproperties	1.0254 (0.000)	1.0353 (0.000)	1.0510 (0.000)
lnwac	0.5165 (0.000)	0.4433 (0.008)	0.3962 (0.030)
lnwedensitywater	-1.9083 (0.000)	-1.8803 (0.000)	-2.3411 (0.000)
lnwedensitywater2	0.1353 (0.000)	0.1339 (0.000)	0.1603 (0.000)
lnboostersperlength	0.3594 (0.001)	0.3369 (0.004)	-
lnaph_ww	-	0.1539 (0.202)	0.1801 (0.117)
constant	-1.9385 (0.069)	-2.9451 (0.020)	-2.7043 (0.143)
R^2	0.9750	0.9763	0.9723
AIC	-140.53	-141.85	-134.54
BIC	-116.03	-114.29	-110.04
RESET	0.0434	0.0390	0.2469
MAPE 2018-20	14.77%	13.97%	14.58%

A.3 Treatment Complexity

Figure A.3.1 – Regression coefficients (p-values) the WW2 specifications with different definitions of weighted average treatment complexity. These use a panel of data from 2011-12 to 2019-20.

Variable	Original WW2 Model	WW2 alt_wac1	WW2 alt_wac2	WW2 alt_wac3
lnproperties	1.0254 (0.000)	1.0331 (0.000)	1.0375 (0.000)	1.0259 (0.000)
lnwac/lnalt_wac1/lnalt_wac2/lnalt_wac3	0.5165 (0.000)	0.6077 (0.000)	0.3694 (0.000)	0.3394 (0.000)
lnwedensitywater	-1.9083 (0.000)	-1.8574 (0.000)	-1.8306 (0.000)	-1.8955 (0.000)
lnwedensitywater2	0.1353 (0.000)	0.1321 (0.000)	0.1298 (0.000)	0.1331 (0.000)
lnboostersperlength	0.3594 (0.001)	0.3790 (0.001)	0.3754 (0.001)	0.3169 (0.000)
constant	-1.9385 (0.070)	-2.1278 (0.070)	-2.4599 (0.042)	-2.3808 (0.001)
R^2	0.9750	0.9735	0.9726	0.9746
AIC	-140.53	-137.01	-135.67	-140.02
BIC	-116.03	-112.50	-111.16	-115.52
RESET	0.0434	0.0403	0.0315	0.0286
MAPE 2018-20	14.77%	15.04%	15.34%	14.97%

Figure A.3.2 – Regression coefficients (p-values) of the WRP specifications with GW and SW treatment split using treatment bands 4-6 and 5-6. These use a panel of data from 2011-12 to 2019-20.

Variable	WRP1 SW and GW Split	WRP2 SW and GW Split	WRP1 SW/GW with SW Control	WRP2 SW/GW with SW Control
lnproperties	1.0334 (0.000)	1.0351 (0.000)	1.0178 (0.000)	1.0147 (0.000)
pctwatertreated46_GW	0.0039 (0.017)	-	0.0036 (0.028)	-
pctwatertreated46_SW	-0.0021 (0.264)	-	-0.0008 (0.678)	-
pctwatertreated56_GW	-	0.0064 (0.006)	-	0.0056 (0.019)
pctwatertreated56_SW	-	-0.0011 (0.344)	-	-0.0008 (0.464)
pctwatertreatedSW	-	-	0.0034 (0.090)	0.0030 (0.172)
lnwedensitywater	-1.6671 (0.001)	-1.3292 (0.042)	-1.6738 (0.000)	-1.4808 (0.018)
lnwedensitywater2	0.1061 (0.002)	0.0870 (0.051)	0.1076 (0.000)	0.0985 (0.017)
constant	-3.9568 (0.016)	-5.4062 (0.012)	-4.0730 (0.001)	-4.8266 (0.016)
R^2	0.9251	0.9008	0.9304	0.9088
AIC	-12.28	-13.49	-14.17	-13.79
BIC	12.22	11.01	13.39	13.77
RESET	0.4292	0.6756	0.7221	0.8214

Figure A.2.3 – Chart displaying groundwater as a proportion of treated water by company. HDD and SVE were combined with DVW and SVT respectively for the purposes of creating this chart.

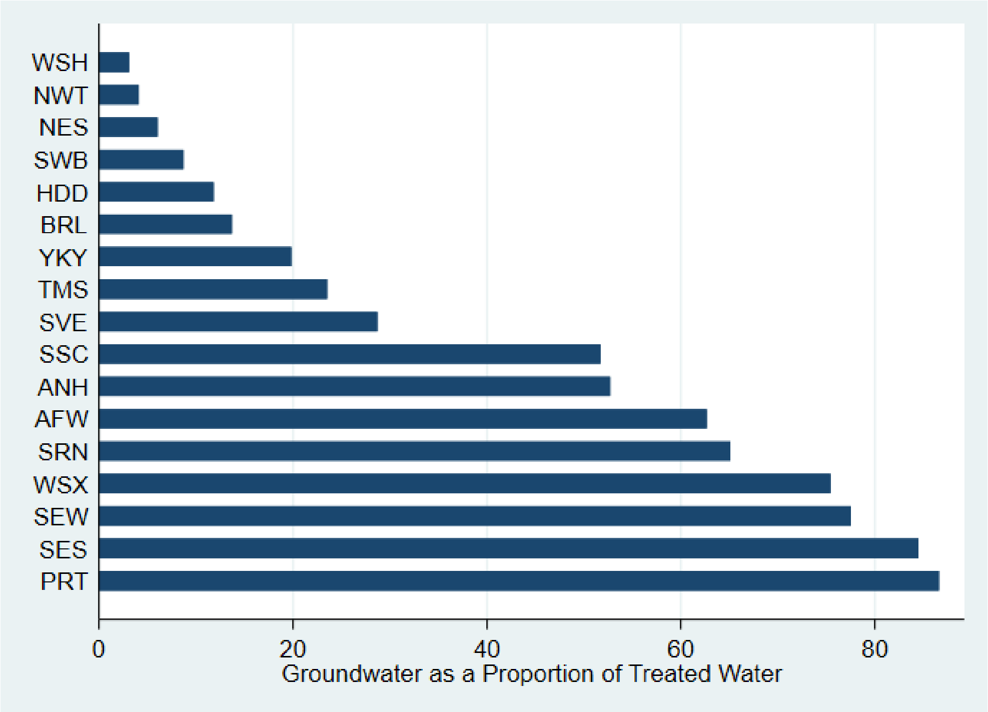


Figure A.2.4 – Chart displaying water treated in complexity bands 3-6 as a proportion of treated water by company. HDD and SVE were combined with DVW and SVT respectively for the purposes of creating this chart.

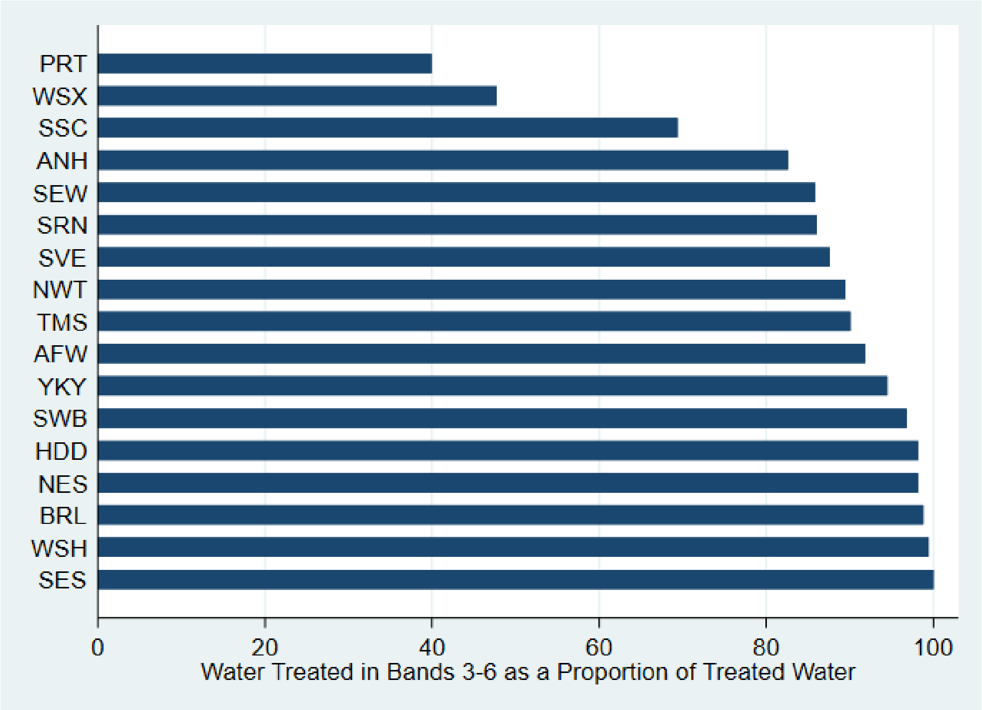


Figure A.2.5 – Chart displaying water treated in complexity bands 5-6 as a proportion of treated water by company. HDD and SVE were combined with DVW and SVT respectively for the purposes of creating this chart.

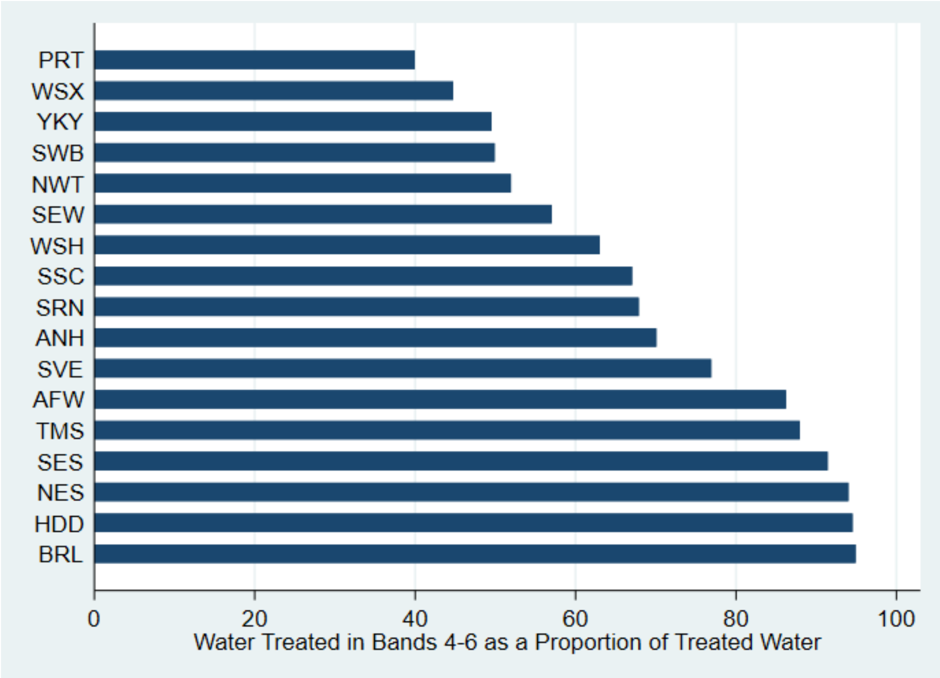


Figure A.2.6 – Chart displaying surface water treated in complexity bands 3-6 as a proportion of treated surface water by company. HDD and SVE were combined with DVW and SVT respectively for the purposes of creating this chart.

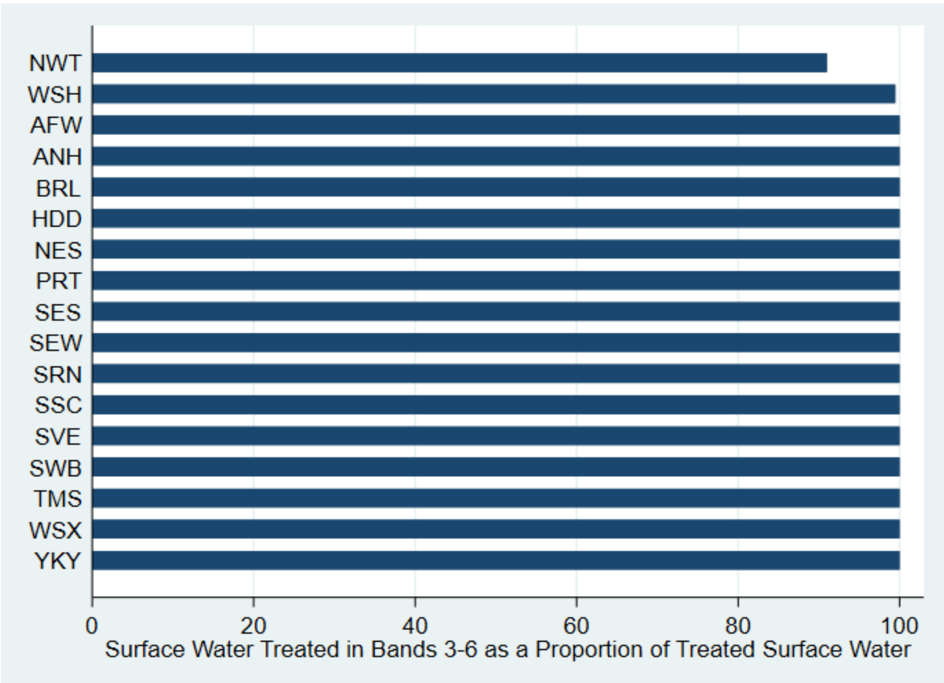


Figure A.2.7 – Chart displaying surface water treated in complexity bands 4-6 as a proportion of treated surface water by company. HDD and SVE were combined with DVW and SVT respectively for the purposes of creating this chart.

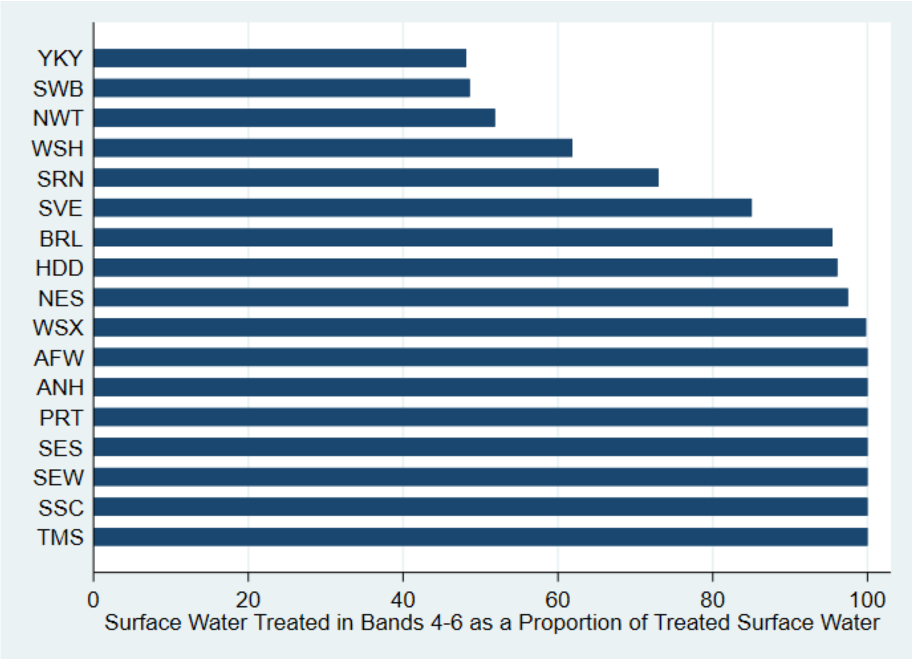


Figure A.2.8 – Chart displaying ground water treated in complexity bands 3-6 as a proportion of treated ground water by company. HDD and SVE were combined with DVW and SVT respectively for the purposes of creating this chart.

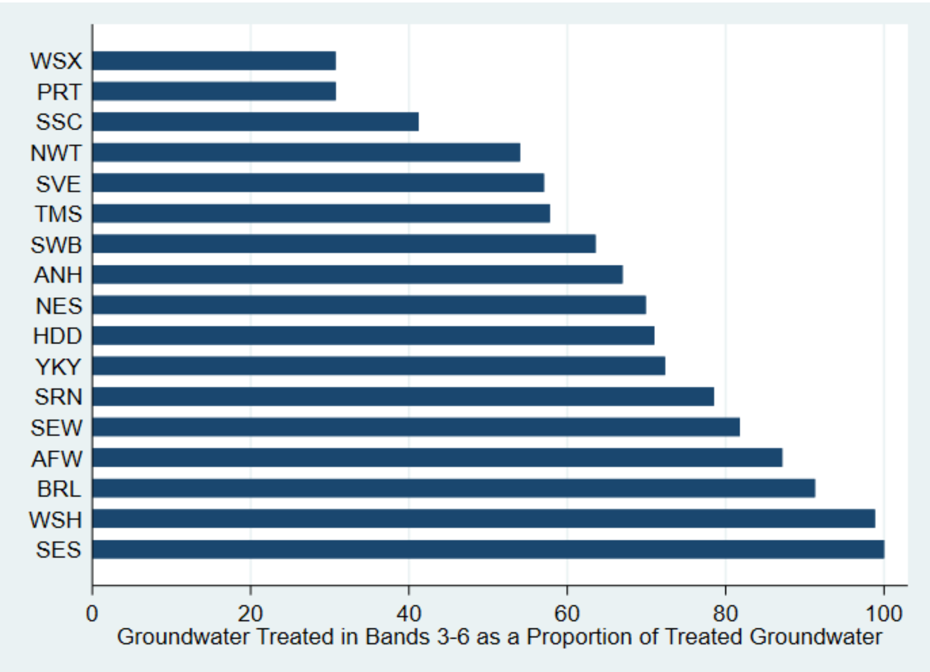
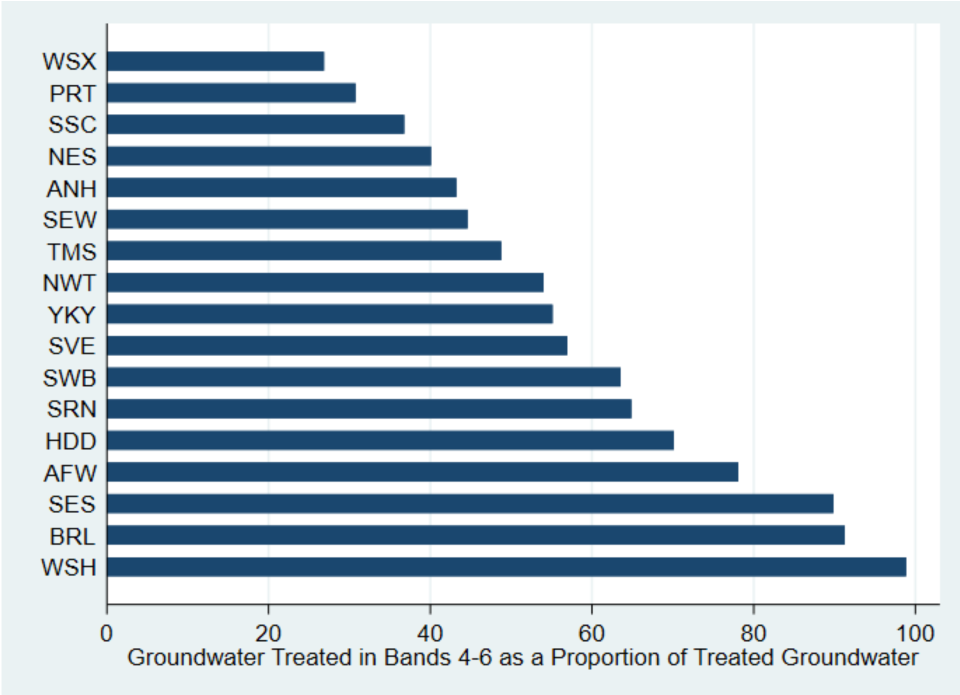


Figure A.2.9 – Chart displaying ground water treated in complexity bands 4-6 as a proportion of treated ground water by company. HDD and SVE were combined with DVW and SVT respectively for the purposes of creating this chart.

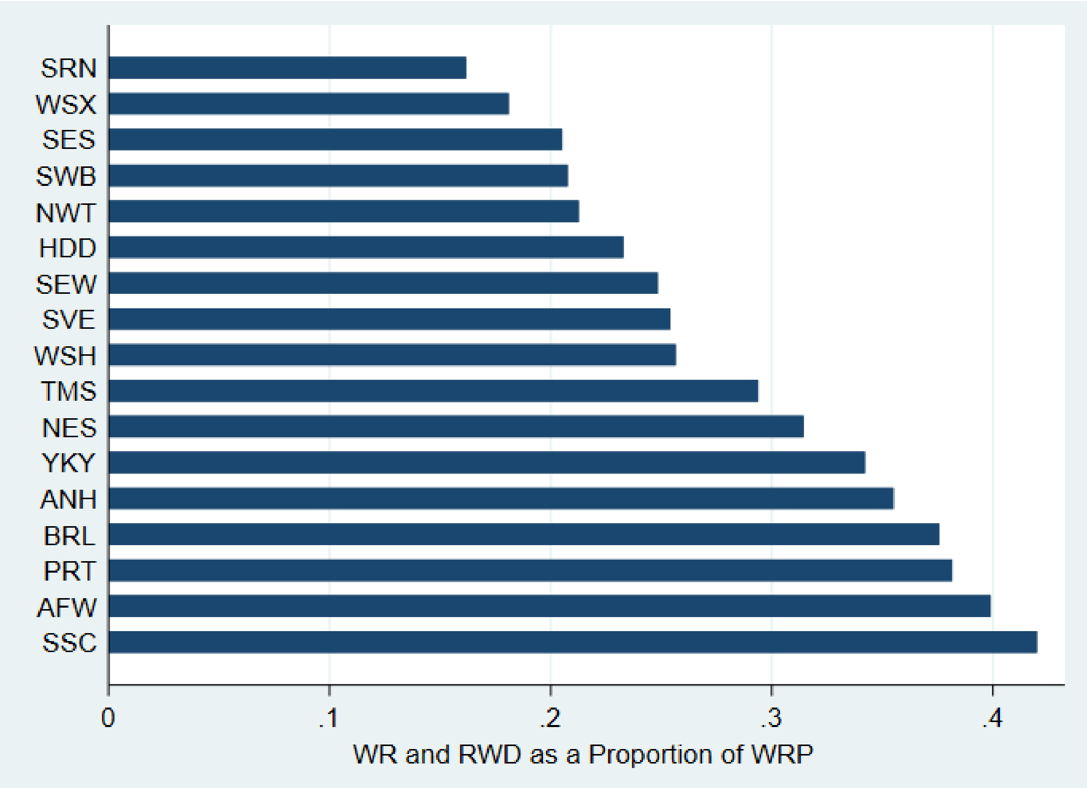


A.4 Water Resources

Figure A.4.1 – Efficiency values based on the water treatment regressions in Figure 3.3.9 relative to the original WRP specifications. Efficiency changes are relative to the original specification.

Company	WRP	Rank	WT	Rank	Efficiency Change
SSC	0.47	1	0.39	1	-17.02%
PRT	0.70	2	0.70	3	0.00%
SEW	0.80	3	0.81	5	1.25%
ANH	0.82	4	0.70	2	-14.63%
AFW	0.89	5	0.72	4	-19.10%
YKY	0.91	6	0.85	6	-6.59%
HDD	0.95	7	0.92	7	-3.16%
NES	0.97	8	0.94	8	-3.09%
TMS	1.05	9	1.05	10	0.00%
SWB	1.09	10	1.26	14	15.60%
WSX	1.10	11	1.19	12	8.18%
BRL	1.15	12	1.04	9	-9.57%
NWT	1.16	13	1.32	15	13.79%
SVE	1.19	14	1.22	13	2.52%
WSH	1.22	15	1.16	11	-4.92%
SRN	1.37	16	1.65	17	20.44%
SES	1.44	17	1.65	16	14.58%

Figure A.4.2 – Water Resources + Raw Water Distribution costs as a proportion of Water Resources Plus costs by company. HDD and SVE were combined with DVW and SVT respectively for the purposes of creating this chart.



A.5 Use of dummy variables for AMPs, AMP Years and Exceptional Years

We believe it essential to make use of indicator variables for AMPs, AMP years and exceptional years.

The Random Effects model can be written as:

$$y_{it} = \alpha + \sum_{j=1}^{p} \beta_j x_{jit} + u_i + \varepsilon_{it}$$

Where y_{it} is the dependent variable for company i at time t, β_j is the coefficient on variable j, x_{jit} is the variable j for company i at time t, u_i is the company specific random effect and ε_{it} is the random error component for company i at time t.

If we add a set of indicators, denoted as π_{it} to the model we can write it as:

$$y_{it} = \alpha + \sum_{j=1}^{p} \beta_j x_{jit} + \sum_{k=1}^{q} \tau_k \pi_{kt} + u_i + \varepsilon_{it}$$

π_{kt} are a set of indicator variables that take the value 1 when some condition is met. They are 0 otherwise. So the coefficient τ_k is added to the model only when the condition is met, and it is not added to the model otherwise.

There will always be one excluded indicator, and the coefficients τ_k can be interpreted as deviations *relative* to that excluded indicator. All future observations will then be treated identically to the excluded indicator.

Accounting for Structural Breaks

We can also consider interaction terms to account for any structural breaks in the data if we believe that the relationship between the explanatory variables and the dependent variable change between AMPs. An output for this can be seen in ***Figure A.5.1***. Given 'amp5' is an indicator variable equal to 1 if the observation is in AMP5, these interaction terms can be interpreted as an adjustment to the coefficient on the variable being interacted with if in amp5. For example, the coefficient on 'lnproperties' is 1.0598 if in AMP5, but 1.0633-0.0635=0.9998 if in AMP5.

Figure A.5.1 – Regression coefficients (p-values) from a modified PR19 WW1 specification including an AMP5 indication and interaction terms.

Variable	WW1 + AMP5 + Interactions
lnproperties	1.0633 (0.000)
pctwatertreated36	0.0081 (0.000)
lnwedensitywater	-2.7253 (0.000)
lnwedensitywater2	0.1912 (0.000)
lnboostersperlength	0.2911 (0.030)
amp5	-2.0193 (0.006)
amp5#lnproperties	-0.0635 (0.037)
amp5#pctwatertreated36	-0.0029 (0.017)
amp5#lnwedensitywater	0.9496 (0.001)
amp5#lnwedensitywater2	-0.0645 (0.001)
amp5#lnboostersperlength	0.0818 (0.558)
constant	0.3026 (0.830)
R^2	0.9758
AIC	-145.81
BIC	-102.94
RESET	0.4664

A.6 Time Trends

Time trends are variables that attempt to capture exogenous trends in the dependent variable that are missed by the independent variables included in the model. If we think that there is something that is exogenously impacting our dependent variable over time, we should include a time trend in the model. Often this is done to account for abstract concepts like 'technological progress' that we can't necessarily find or create a single neat variable for.

What are Time Trends in the Regression Context?

Again, we can write our random effects model as in ***Appendix A.5***.

If we add a time trend to the model, we can write it as:

$$y_{it} = \alpha + \sum_{j=1}^{p} \beta_j x_{jit} + \lambda t + u_i + \varepsilon_{it}$$

The t denoting the time trend is the same as the t in y_{it} etc., so y_{i1} would mean that we add $\lambda \times 1$ to our model, y_{i2} would mean we add $\lambda \times 2$ and so on.

Differentiating y_{it} with respect to t gives us:

$$\frac{dy_{it}}{dt} = \lambda$$

This means that a unit increase in t increases y_{it} by λ and this is the same for all time periods.

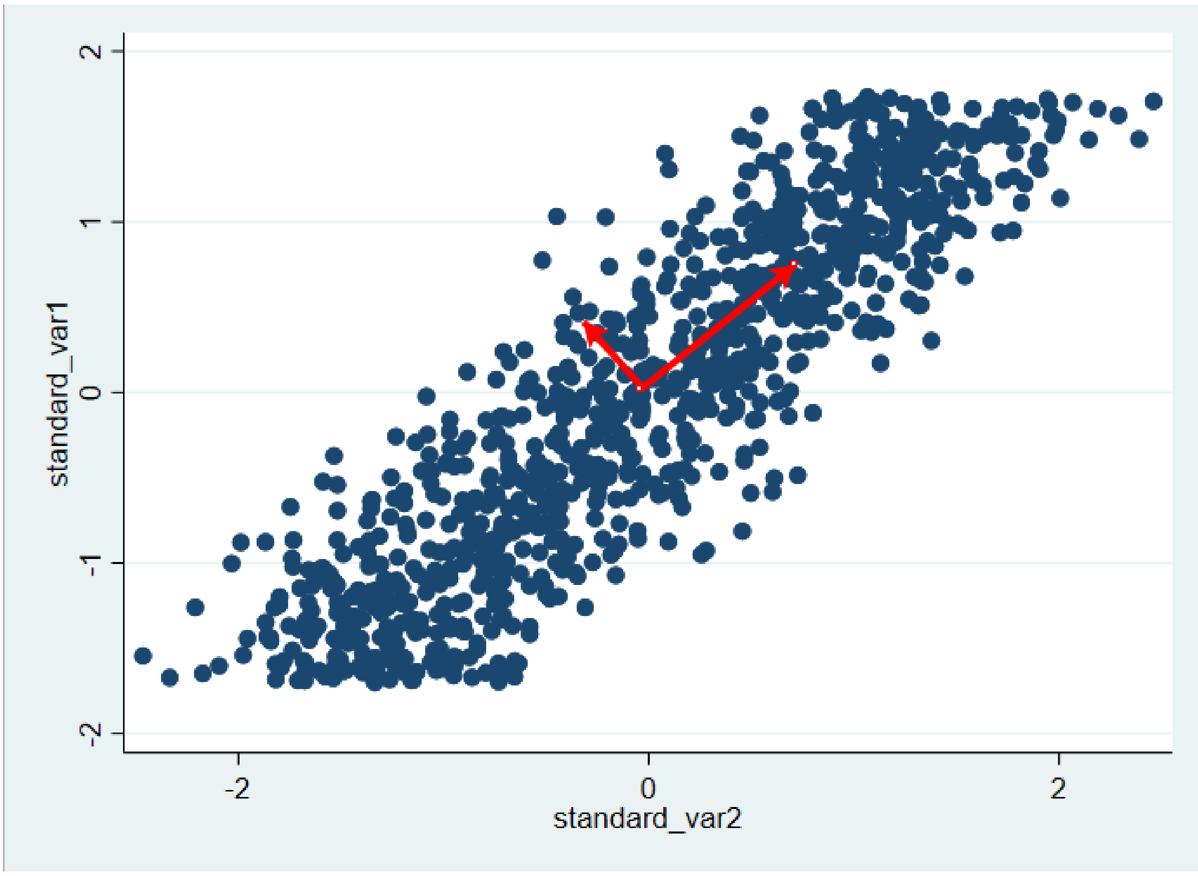
A.7 Principal Component Analysis

What is Principal Component Analysis (PCA)?

Principal Component Analysis is a dimensionality reduction technique that reduces many variables into fewer variables while preserving most of the information from the original set. The reduced set of variables will be orthogonal by design, so will not suffer from any problems associated with multicollinearity. As such we can also use this technique to reduce a set of highly correlated variables that would suffer from multicollinearity in a regression, into a smaller set of variables that work well in the regression context. A good example would be a set of scale variables.

A Graphical Representation

Figure A.7.1 – A scatter plot showing two randomly produced and standardised variables and two principal components in red, the first on the right and the second on the left.



The first principal component is the vector that explains the highest amount of variance in our data. Further principal components are the vectors that explain the highest amount of variance while being orthogonal (at a right angle) to all the previous principal components.

As seen in the graph above, our first principal component - the red arrow to the right - is in the direction of the most variance in the data, i.e. the direction in which the distribution is widest. The second principal component is then in the direction of the 'next most' variance in the data *and* orthogonal to the first principal component.

Deriving the Principal Components

The first step in deriving the principal components is to standardise our variables, x_j. For each j, we can write the mean as μ_j and the standard deviation as σ_j. We can then standardise our variables using:

$$\tilde{x}_j = \frac{x_j - \mu_j}{\sigma_j}$$

This means that all the variables will be of the same scale and therefore the variance maximisation problem will not be disturbed by 'larger' variables with greater *absolute* variation erroneously suggesting greater variance. It also 'de-units' the data, so we remedy any issues around different variables being expressed in different units.

We then compute the covariance matrix for our variables. For two variables, x and y, we can write:

$$cov(x, y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Denoting this as Σ_{xy}, we can then write the covariance matrix for three variables x, y, z as:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} & \Sigma_{xz} \\ \Sigma_{yx} & \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zx} & \Sigma_{zy} & \Sigma_{zz} \end{bmatrix}$$

We then solve for the eigenvalues of $\boldsymbol{\Sigma}$, which are the values of λ for which the following holds true:

$$det(\boldsymbol{\Sigma} - \lambda\boldsymbol{I}) = 0$$

$\boldsymbol{I}$ is the identity matrix, a matrix which has all elements equal to 0 other than the primary diagonal for which all elements are equal to 1, so it's is equivalent to solving for λ in:

$$det\left(\begin{bmatrix} \Sigma_{xx} - \lambda & \Sigma_{xy} & \Sigma_{xz} \\ \Sigma_{yx} & \Sigma_{yy} - \lambda & \Sigma_{yz} \\ \Sigma_{zx} & \Sigma_{zy} & \Sigma_{zz} - \lambda \end{bmatrix}\right) = 0$$

det is the 'determinant' of the matrix, which gives us the eigenvalues of $\boldsymbol{\Sigma}$. We can then solve for the corresponding eigenvectors, which are the vectors $\mathbf{v_j}$ for which:

$$\boldsymbol{\Sigma}\mathbf{v}_{\boldsymbol{j}} = \lambda_j\mathbf{v}_{\boldsymbol{j}}$$

i.e. an eigenvector is one that only changes by a scalar factor (the eigenvalue) when a linear transformation ($\boldsymbol{\Sigma}$) is applied. It can then be rewritten as:

$$(\boldsymbol{\Sigma} - \lambda_j\boldsymbol{I})\mathbf{v}_{\boldsymbol{j}} = \mathbf{0}$$

There will be an eigenvector for each eigenvalue if the determinant of the matrix can be found. The $\mathbf{v_j}$ corresponding the n^{th} largest eigenvalue will be the n^{th} principal component. So to find the first principal component, we select the largest eigenvalue, λ_1, and solve for $\mathbf{v}_1$.

We can then create a matrix $\boldsymbol{W}$, a $p \times n$ matrix where p is the number of principal components we wish to use, and n is the number of variables included in the analysis. We the create our principal components, $\boldsymbol{k}$, using:

$$\boldsymbol{k} = \boldsymbol{W}'\boldsymbol{x}$$

$\boldsymbol{x}$ is the vector for each observation of our n variables included in the PCA, and $\boldsymbol{W}'$ is the transpose of the matrix $\boldsymbol{W}$.

Selecting the Number of Principal Components

The eigenvalues also give us information on how many principal components we should be using. The proportion of variance explained by any principal component can be calculated as:

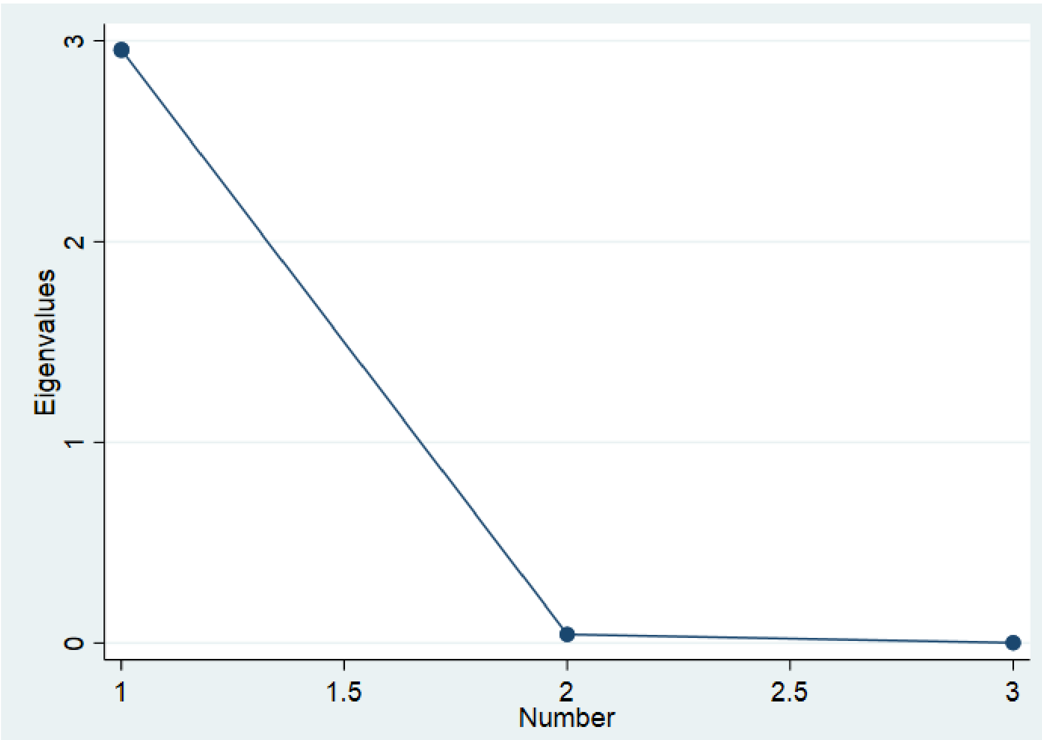
$$\frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}$$

The ratio of the i^{th} eigenvalue to the sum of all the eigenvalues is the proportion of variance explained. We can then make the decision to include only principal components that explain over a certain amount of variance or eigenvalues over a certain size.

We can also generate a 'scree plot' that graphs the eigenvalues so we can see visually the proportion of variance explained.

There is no hard and fast rule here, but we might only want to include components that explain more than 5% of the variance or that have an eigenvalue greater than 1 for example.

Figure A.7.2 – A scree plot of eigenvalues following PCA. It serves as a visual representation of the proportion of variance explained by each component. In this case the first principal component clearly explains the vast majority of the variance and will likely be the only one included in any further analysis.



Interpretability of the principal components

The main concern we might have with PCA is that it renders the variables and in particular their coefficients less interpretable. But the information will be preserved and the association will still be present, just 'hidden' behind the principal components. Since a principal component is just a weighted average of the standardised variables, we can feasibly decompose the principal component back into the original variables and find the implied coefficient attributable to each of the original variables.

For example, for the 'pca_scale' driver used for the regression in ***Figure 3.3.1***, the first principal component uses weightings of 0.5802, 0.5733 and 0.5785 for log properties, log lengths of main and log population respectively. Multiplying the coefficient on 'pca_scale' by each of these weightings will give the implied coefficient on the respective standardised scale variable.

Reducing Large Specifications into Principal Components

We could potentially go as far as reducing a very large set of explanatory variables down to several principal components and using those components as the explanatory variables in regression analysis. This would allow us to generate a lot of information for our models using an appropriate number of variables.

If we select all the variables that we might suppose have a material effect on the respective base cost because of engineering logic, we can reduce them to smaller set of principal components and regress the relevant base costs on the principal components.

This might be valuable in the NPWW specifications, where variables often have an opposing effect on the SWC and SWT components of cost.

Composite Variables

A more reasonable use for PCA might be to produce a composite of variables we know are strongly correlated but that each have their own engineering rationale for inclusion in the model. Consider the scale variables in treated water distribution. The length of mains is obviously associated with costs, the more mains a company has, the more costs are incurred by replacements and maintenance along those mains. Properties is also associated with costs though, the more properties a company has, the more water will be moving through the mains, and the more joins there are from the mains – increased usage increases pressure which increases bursts, particularly at joins which are the weak points in the network. Population is a further indicator of water usage, with different companies perhaps having different numbers of people in a household there is an argument that this should also be included.

Previously, we have selected a single scale variable to include in each model. Scale variables will always be highly correlated, with large companies having large populations and large networks. However, there may be minor variation in these, with one company having more mains than another but fewer properties. Which of these is the best indicator for each of the specifications is debatable, so we can use PCA to create a composite variable which can tell us how large companies are overall.

Computing Future Predictions

The final consideration is the prediction of the future values. We can't just forecast forward the PCA. Instead, we must forecast forward all of the variables included in the principal component analysis, then apply the formula $\boldsymbol{k} = \boldsymbol{W}'\boldsymbol{x}$ to the forecast variables, i.e. we find our $\boldsymbol{W}'$ from the original principal component analysis, then derive our principal components using it.

A.8 AIC, BIC, RESET and VIF

Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC) – testing for overfitting

The Akaike Information Criterion uses the generated log likelihood of a model but also penalises based on the number of coefficients being predicted. This provides valuable information that can be used to help identify models that fit the data well while avoiding the problems associated with overfitting. It is calculated as:

$$AIC = 2k - 2\ell(\theta)$$

$\ell(\theta)$ is defined as the log likelihood of model θ and k is the number of estimated coefficients in the model specification. The best quality model is the one that minimises the AIC. This is a model that explains the greatest amount of variation using the fewest possible independent variables and thus limits the risk of overfitting.

A corrected version of AIC (AICc) can also be used that weights the penalty for adding parameters in small samples where there is an increased likelihood of overfitting. It is calculated as follows:

$$AICc = \frac{2kn}{n-k-1} - 2\ell(\theta)$$

Here n is the number of observations used. Given the relatively small size of the data panel, this is likely to be useful if considering the specification of models of different panel lengths.

The Bayes Information Criterion is very similar to the AIC but takes greater account of both the sample size and the number of parameters used. This becomes more punitive if a larger sample size has been required to derive an equivalent likelihood factor. It can be calculated as:

$$BIC = ln(n)k - 2\ell(\theta)$$

The parameters here are defined in the same way as in the AIC explanation. Again, a model with a lower BIC can be considered 'better'.

We do not think that this type of information should be used to systematically remove cost drivers that can be justified through engineering logic and empirical evidence. However, it will provide an appropriate prompt to review and ensure that the underpinning case is secure in cases where the AIC or BIC values increase when the further explanatory variables are added. The information can also be used comparatively to identify preferred variables for a specific cost driver being included in the model.

RESET Test – testing for misspecification

The RESET test looks for model misspecification by adding the squared, cubic and potentially higher order terms of the normalised fitted values from the original linear regression. This is essentially adding non-linear terms to the model to see if that have predictive power. Given the potential for overfitting and the ability to interpret the results in the observed data, we suggest that it might be appropriate and pragmatic to limit the test to quadratic or potentially cubic extensions rather than continuing to higher orders.

The joint significance of the extra terms is tested. Where they are jointly significant, the null hypothesis that there is no misspecification is rejected. We do not think that it would be appropriate to systematically reject models or add quadratic terms using this information alone. Instead, a RESET test failure should be a trigger the revisiting of engineering logic and empirical evidence to see if there is justification for including targeted non-linear explanatory variables or interaction terms. This means that the primacy of engineering logic will be retained.

Variance Inflation Factor (VIF) – testing for multicollinearity

The variance inflation factor is a simple test used to assess multicollinearity in a regression model. It identifies the strength of correlation among the predictors. This is done by sequentially assessing the R^2 of each individual explanatory factor against all the others in a model. The formula is given below, where j is an indicator for each variable in the model, and R_j^2 is the R^2 from a regression of the j^{th} dependent variable on the other dependent variable:

$$VIF_j = \frac{1}{1 - R_j{}^2}$$

It is sensible to review the VIF values in a model. Where values are above predetermined thresholds (e.g. 5), further consideration might be appropriate. This could be removal of variables where the identified explanatory drivers are both likely to drive costs in a linear fashion, or transformation of a variable where a different explanatory power is anticipated.

However, high VIF terms are expected where models include a variable and its squared term or interaction terms. There is no reason to exclude these variables where the functional forms can be justified. It may also be appropriate to ignore high VIF values where the identified variables are not of particular interest in themselves, or where dummy variables have been included that represent multiple categories.