Wessex Water
YTL GROUP

Direct line: [redacted]
Email: [redacted]

12 May 2023

Dear Cost Assessment team,

Econometric base cost models for PR24 – Wessex Water response

Thank you for the opportunity to respond to the consultation on Econometric base cost models for PR24. Our consultation response comprises this letter and the enclosed appendix, taken together.

1. Progress achieved since PR19 and further opportunities ahead of draft determinations

We believe that the econometric models used for price control cost assessment should be subject to a process of continual improvement across successive price reviews. We welcome endeavours to increase the extent to which the inevitably simplistic econometric models capture the underlying cost drivers that lead to differences in the levels of efficient costs between companies.

It is clear that Ofwat has put considerable effort into the development of its set of consultation models. It has rightly taken the PR19 models as a starting point, but it has not been unduly constrained by these models. We can see this in the expansion of the cost drivers covered (e.g. to capture urban rainfall), in the further development of the types of variables used at PR19 (e.g. density), as well as in the proposed move to unit cost models for bioresources.

In addition, we found that Ofwat's consultation document was well-written and well-organised, complemented by useful supporting information (e.g. Stata code, updated data, and the spreadsheet with model estimation results laid out clearly).

We have developed a detailed response to Ofwat's consultation, which we hope that Ofwat finds to be both constructive and interesting.

We have identified some serious shortcomings and missed opportunities with aspects of Ofwat's modelling as it stands. To a large degree this reflects the complexity of the subject matter, and shows the value of an iterative process of model development, exposition, review and refinement.

In this context, we consider it highly important that there is a plan for further engagement with stakeholders on these models – which have such a large impact on customer bills – before Ofwat publishes draft determinations in over 12 months' time.

In particular, we suggest that Ofwat releases an updated and refined set of models in autumn 2023 (around the same time that companies submit business plans). This would allow for review of models in the light of an additional year of data and for Ofwat to show the modelling refinements it has implemented in the light of the stakeholder feedback on the current consultation.

We would also be keen to engage directly with Ofwat in the interim on specific issues raised by our response, and to help ensure that Ofwat has access to the full set of evidence that it needs.

2. Points of substance

We provide a detailed response to Ofwat's consultation models and consultation questions in the appendix to this letter enclosed. Given the number of models and questions that Ofwat is consulting on, and the scope of the model specification options available, we have not sought to summarise all our comments here. We highlight some selected points below:

- **The evidence used for model development and selection**

  Our view is that Ofwat's econometric benchmarking activities are held back by limitations in the types of information used to inform model assessment and selection. This primarily concerns the various econometric results and tests that Ofwat uses to address its cost assessment principle of "robust econometric cost models". We see the key opportunities for improvement in Ofwat's assessment of model robustness to come from enhanced metrics that provide highly relevant information on the *relative performance* of alternative models. In particular, concerning:

  (a) the sensitivity of modelled costs to the impact of dropping companies/years from the sample; and
  (b) a normalised measure of goodness of fit which does not suffer from the limitations of R-squared of the spread of efficiency ratios.

  We found these metrics to have been very useful in reviewing Ofwat's consultation models. Once it becomes familiar with them, we would expect Ofwat to come to value having these metrics at its disposal for model development and selection as it proceeds through the process towards draft and final determinations.

- **Time trend variables**

  There is considerable evidence that, over the timeframe covered by Ofwat's models, industry-wide costs (relative to CPIH) have changed considerably, in ways that are not explained by movements in the cost driver variables captured in Ofwat's econometric models. None of Ofwat's consultation models have a time trend variable and Ofwat rejects the use of time trend variables as a matter of principle. However, Ofwat's position on the use of time trends in its econometric benchmarking models does not seem to be sustainable. It is not evidence-based, and it conflicts with standard statistical / econometric practice for time series panel data. Even if Ofwat thinks that future cost trends will differ from those in the past, this is something that can be addressed in the way that allowances for the 2025-30 period are calculated, even if a time trend is used to model the past.

Our position is not that all parts of the value chain warrant the inclusion of a time trend variable, but that this should not be dismissed as a matter of principle and should instead be informed by case-specific evidence from estimation of models with time trends.

- **Exogenous versus endogenous variables and sewage treatment**

  In line with our broader support for outcomes-based regulation, we consider that Ofwat's cost benchmarking models should be making as much use as possible of explanatory variables that are truly exogenous rather than variables that are reflective of the inputs used by a company and its past management decisions. Ofwat's consultation models contain a mix of explanatory variables, some of which are truly exogenous and some endogenous (i.e. under some degree of management influence over time). We see a legitimate role for endogenous variables as a fall-back option where the underlying cost drivers are difficult to capture well via exogenous variables. But even in these cases, it makes sense to use models with endogenous variables alongside, rather than in preference to, models with exogenous variables.

  In this context, we propose that Ofwat expands its set of SWT models to include some with exogenous density variables. Each of its current models for sewage treatment involve variables relating to treatment works size that are not exogenous. In terms of the underlying cost drivers affecting the *efficient levels* of sewage treatment costs, what matters is the *opportunities for economies of scale* rather than the actual size of sewage treatment works; exogenous measures of population density are highly relevant to the differences between companies in these opportunities.

- **New density metrics and water resources plus**

  Ofwat's consultation models include two new population density metrics, calculated using MSOA-level data. Our view is that there is not a good basis for forming an *a priori* position on which of these is the most appropriate to use to capture density in the econometric models. This is best approached as an empirical matter and the choice of metric(s) may differ between different parts of the value chain.

  Drawing on analysis across the different model variants that Ofwat is consulting on, we found that for water resources plus, the models with the "Weighted average density – MSOA" variable performed considerably worse statistically, especially in terms of the sensitivity of modelled costs to dataset variations. For treated water distribution and sewage collection, the results were more mixed, suggesting a potential role for both metrics within the suite used to set allowances.

Please see the appendix for our full set of comments, which cover a wider set of issues.

Wessex Water
Claverton Down
Bath BA2 7WW
Tel 01225 526 000
Web wessexwater.co.uk

Wessex Water Services Limited registered office
Registered in England No 2366648

3. Scope of the consultation

Ofwat's consultation is quite narrowly focused on one part of its overall process for using econometric models to determine allowances for base-plus costs: *the specification of individual econometric models for different parts of the value chain*. Our response has a corresponding focus.

Nonetheless, we see value in significant further engagement, well before draft determinations, on a number of methodological issues that relate to the way that the econometric models are used. For instance:

- **Scope and diversity in the model suite**

  There are choices to make on how many models to include for each part of the value chain within the suite of models to be used to set allowances (as a means to "mitigate the risk of error and bias in any one model" as Ofwat explains in its consultation). Besides the number of models to be used, there are choices for the composition of the model suite which affect the extent of diversity across reasonable modelling approaches or variables.

- **Aggregated versus disaggregated models**

  There are choices to make about the weight, in the triangulation across the suite of models used to set allowances, between combinations of disaggregated models versus aggregated models. The types of enhanced metrics on modelled cost sensitivity and goodness of fit highlighted above can help provide evidence to inform these decisions (the conventional econometric results/tests are not well-suited to comparing the performance of one group/suite of models against another).

- **Off-model adjustments for growth-related and drainage-related enhancements**

  Given the inclusion of network reinforcement, growth-related enhancements and enhancements relating to drainage and flooding risk in Ofwat's base cost models, there will be a need for some form of off-model adjustment for differences between companies, and over time. While this can draw on the approach from PR19, we see a need to consider improvements, especially in relation to drainage- and flooding-related expenditure that is not simply driven by population growth.

- **Catch-up efficiency adjustments**

  There are questions about the extent to which differences between modelled costs and actual costs are reflective of efficiency differences, which has implications for the adjustments made to reflect the level of costs of a notional efficient company. These questions are particularly relevant given the scale of differences across different parts of the value chain in the efficiency scores arising from Ofwat's models (e.g. a range of around 70% to 150% for wholesale water compared to 90% to 110% for wastewater network plus).

As stated previously, we also have concerns about the separation of benchmarking and efficiency challenges between bioresources and wastewater network plus. For cost assessment to be outcomes-focused and coherent, and for an appropriate efficiency challenge to be set in the round across price controls, attention is needed to cost interactions between bioresources and wastewater network plus. This attention is needed to mitigate the risk of setting an infeasible efficiency challenge overall, and to keep the link between cost and service in focus.

In addition, we would welcome the opportunity to engage on econometric models for enhancements that are modelled separately from base expenditure.

4. The structure of our response

We have provided some quite detailed and constructive comments in the appendix, alongside supporting evidence. The appendix is organised as follows:

- **A1: The evidence used for model development and selection.**
  This provides proposals for improvements in the types of metrics/evidence used by Ofwat to assess the robustness of candidate econometric benchmarking models.
- **A2: Changes in efficient costs over time and time trend variables**.
  This responds to Ofwat's consultation position on the use of time trend variables, which is a cross-cutting issue applicable to wholesale and retail cost assessment.
- **A3: Our response to the consultation questions**.
  This takes each of Ofwat's questions in turn and provides our response. We have submitted alongside this document the Excel submission template, however our full responses to the consultation questions are enclosed in Appendix A3. Due to inclusion of graphs and references to external reports, we have found it difficult to convey our response in a comprehensive way in the excel submission.

In addition, some parts of our appendix refer to work we commissioned from Reckon on residential retail cost assessment and we have provided URLs to the relevant reports within the appendix. We have also submitted these alongside our consultation response for completeness.

5. Next steps

I hope you find this response helpful and would welcome further discussions on any of these points as we look to work collaboratively to deliver an improved suite of econometric models for PR24.

Kind regards,

[redacted]

Matt Greenfield

Director of Strategy and Regulation

Wessex Water
Claverton Down
Bath BA2 7WW
Tel 01225 526 000
Web wessexwater.co.uk

Wessex Water Services Limited registered office
Registered in England No 2366648

A1: The evidence used for model development and selection

Ofwat's consultation includes a variety of econometric models for different parts of the value chain, including models with new or refined explanatory variables that were not available for PR19. An immediate question is which of these models – as well as possible further model variations – are the most appropriate to use for the purposes of cost assessment at PR24.

However, there is a preliminary question, which deserves much greater attention than it seems to have had so far in the PR24 process: *what techniques and evidence are most appropriate to use for the purposes of deciding which models to use for cost assessment at PR24?*

If the wrong tools and techniques are used to select between candidate models, it is difficult to have confidence that the models selected will be appropriate. Neither Ofwat's current consultation, nor its earlier consultations in the PR24 process, have properly engaged with this question.

Our view is that Ofwat's econometric benchmarking activities are held back by limitations in the types of information used to inform model assessment and selection. This primarily concerns the various econometric results and tests that Ofwat uses to address its cost assessment principle of "robust econometric cost models". This issue is not isolated to Ofwat; looking across the industry, the techniques used for model development and selection are not as good as they could be – or indeed should be – especially given the scale of customer funding riding on them.

We see significant opportunities for improvement for the remainder of the PR24 process. We commented on the topic briefly as part of our model submissions to Ofwat in January 2023. We expand on this point below, taking the following topics in turn:

- The objective of the econometric model development process.
- The limitations of the types of tests used by Ofwat.
- Key areas for improvement in the assessment of model robustness.
- Sensitivity of model estimation results to changes in the assumptions and data.
- A more usable and useful metric of goodness of fit.
- Examples of practical benefits.

The objective of the econometric model development process

Ofwat explains its view that its cost models "should accurately predict and forecast efficient costs" and that, to achieve this, it has assessed the models against a range of model robustness tests.[1]

However, the objective that the econometric cost models should accurately predict and forecast efficient costs is not realistic. It is not realistic to expect a set of econometric models of the type used by Ofwat to *fully capture* differences between companies that affect their efficient levels of costs (or unit costs), such as differences in their operating environments and the services they need to provide to customers. Ofwat recognises this when it says: "We triangulate across a set of models with different cost drivers and levels of cost aggregation to mitigate the risk of error and bias in any one model".

[1] Ofwat (2023) *Econometric base cost models for PR24*, page 16.

It is more realistic and helpful to see the task as that of developing econometric models that, **as far as practical**, accurately predict and forecast efficient costs. This perspective recognises that, even with the best possible model development and assessment process for PR24, the models produced will not accurately predict and forecast efficient costs.

Crucially, this alternative perspective highlights that what matters is not the *absolute performance* of any specific model ("is it robust?"), but rather the *relative performance* of alternative feasible models. This has implications for the types of evidence used for assessing model robustness.

The limitations of the types of tests used by Ofwat

Ofwat's approach to the assessment of model robustness has no shortage of tests. Leaving aside the information reported for each model on the p-values and purported statistical significance of individual coefficients (which we agree is of high importance), Ofwat's consultation reports, for each model, 17 different tests or metrics relating to model robustness.

However, this approach has two major shortcomings:

1. It gives too much emphasis to pass/fail tests that are not particularly useful for the purposes of assessing the relative performance of alternative models under consideration.
2. It misses out on good ways to capture the most important aspects of model robustness and model performance.

We focus briefly in this subsection on the first point above before proceeding to the second in the subsections that follow.

Most of the diagnostic tests reported in Ofwat's consultation do not actually provide a good way to judge the relative merits of alternative models. We give an example of this in the box below in relation to the RESET test.

Example of diagnostic test limitations for model selection purposes: RESET test.

We take the example of the RESET test (regression specification error test) and consider what a positive test result actually tells us.

Suppose we find that consultation model A fails the RESET test and that consultation model B passes this test, at a given confidence level. These results might be indicative of potential opportunities for finding a variant on model A, with a more complex functional form (perhaps involving quadratic or interaction terms) that could fit the data better than model A. This finding might be useful as a means to discover candidate model C as a potential improvement on model A.

But this is of very limited relevance to the choice between models A and B in circumstances where there are differences between these models in the set of explanatory variables (not just functional form). In logical terms, a test result that there may be a model C that outperforms model A does not provide any sound basis to favour a separate model B over model A.

If an assessment is being made between models A and B, then – for the purposes that Ofwat uses benchmarking models and the types of models being compared – it would generally be an error to give any weight to the fact that model A failed the RESET test and model B did not.

Our reading of the consultation is that Ofwat might not have made much use of its diagnostic tests. For instance, while Ofwat highlights the role of the RESET test within its approach to assessing model robustness,[2] the only practical references in the consultation document to the RESET are cases where Ofwat states that it found certain models failed the RESET test but this was not actually a concern.[3] However, it is possible that some otherwise promising models or approaches have been lost due to misplaced concerns arising from diagnostic tests.

In any event, despite the array of model robustness tests reported, Ofwat has had to make do with rather limited information on which to compare the performance and robustness of alternative models.

Rather than pass/fail tests or red/amber/green ratings, it would be much better to use *metrics* that capture a model's statistical performance, and which can be used to compare the performance of alternative models in *relative terms.*

Some of Ofwat's reported results on model robustness/performance do enable direct comparisons of relative performance across models (in some cases at least). For instance, this is a feature of the range of implied efficiency scores across models for the same part of the value chain. Ofwat quite reasonably states that "A large range of efficiency scores could indicate the presence of issues in the underlying model, such as the presence of omitted variables".[4] However, Ofwat does not seem to mention this metric at all in the interpretation or assessment of model estimation results in the main sections of its consultation.[5] Ofwat is perhaps concerned about the limitations of a metric that is dependent only the most extreme observations in the dataset.

Key areas for improvement in the assessment of model robustness

We see the key opportunities for improvement in Ofwat's assessment of model robustness as follows:

- **Sensitivity of modelled costs**. A key aspect of model robustness is the sensitivity of the *modelled costs* derived from a specific model to variations in the dataset (e.g. dropping companies and years). The modelled costs are the key results that Ofwat takes from its econometric models and uses to set allowances, and these reflect the combined effect of the estimated coefficients on individual explanatory variables. Ofwat's consultation does use some sensitivity analysis, but this is not focused on modelled costs, and it is of very limited scope.

- **Goodness of fit**. Ofwat correctly identifies that the R-squared measures should only be used to compare models with the same dependent variable. There are other limitations of R-squared. Nonetheless, goodness of fit – as a broader concept – is a highly relevant aspect of model performance. We see considerable value in a *normalised* metric of goodness of fit that can be used to compare models with different dependent variables and to compare the goodness of fit between alternative suites of models (e.g. suites based on aggregated models versus suites built up from disaggregated models).

We provide further information on these opportunities for improvement below.

---

[2] Ofwat (2023) *Econometric base cost models for PR24*, page 16.
[3] Ofwat (2023) *Econometric base cost models for PR24*, page 56 and 66.
[4] Ofwat (2023) *Econometric base cost models for PR24*, page 71.
[5] Other than a brief reference (page 33) to why it thought that an increase in the range of efficiency scores in its consultation models compared to the position at PR19 was not an indication of any worsening in its models since PR19.

Importantly, we see these as *building on* and improving concepts, tests and metrics that Ofwat already uses and the underlying considerations that motivate aspects of the approach it used for the consultation (e.g. sensitivity/stability tests). Ofwat is right to be concerned with the sensitivity of model estimation results; it just needs a better way to assess this. Ofwat is also right to be concerned with goodness of fit; it just needs a better metric.

Sensitivity of model estimation results to changes in the assumptions and data

Ofwat rightly recognises that an important aspect of model assessment concerns the stability of model estimation results to changes in the underlying data.

Ofwat frames this is a binary yes/no test:[6] *Are the estimated model results stable / robust to changes in the underlying assumptions and data*?

However, the framing of the question in this way is not helpful for practical purposes. For the types of models being considered for cost benchmarking purposes, model estimation results will inevitably *change to some degree* as a consequence of variations in the assumptions and data. We see no sound basis for setting a threshold as to what degree of instability counts as too much for the model to be deemed robust or acceptable. Even if there were to be such a threshold, this would be unhelpful for model selection purposes in cases where multiple models fall short of that threshold or where multiple models exceed that threshold.

It would be much more useful to instead frame this question in relative terms: *to what extent are the model results stable / robust to changes in the underlying assumptions and data?*

Furthermore, it is highly important to properly consider which aspects of "model results" such stability is most important for. Ofwat's sensitivity tests were defined as follows:[7]

> *"This is a test to assess robustness of the model to changes in the underlying assumptions. Robustness under the first test should be assessed by removing the most efficient company, and separately the least efficiency company from the sample. Robustness under the second test should be assessed by removing the first year of the sample, and separately the last year of the sample. Results of the test should be reported using the following RAG rating (the lower the rating, the less confident we are in model stability): Red (R): the estimated coefficients present changes in both significance and sign; Amber (A): the estimated coefficients present some changes in significance but not in sign; and Green (G): the estimated coefficients do not present changes in significance or sign."*

This approach focuses on whether four specific variations in the dataset (dropping first last / year and company with the lowest/highest efficiency score) has effects on (a) the sign (i.e. positive or negative) of coefficients and (b) the purported level of statistical significance of those coefficients (i.e. p-values changing between the four categories of significant at 1%, 5%, 10%, or none of these). It is difficult to understand the basis for testing sensitivity in this way. Even if the focus is to be on the sensitivity of estimated coefficients, it seems odd to give no regard whatsoever to changes in the magnitude of estimated coefficients that remain of the same sign.

---

[6] Ofwat (2023) *Econometric base cost models for PR24*, page 16.
[7] Ofwat (2022) *Template and guidance for the submission of base econometric cost models ahead of the spring 2023 consultation*, page 18.

Wessex Water
Claverton Down
Bath BA2 7WW
Tel 01225 526 000
Web wessexwater.co.uk

Wessex Water Services Limited registered office
Registered in England No 2366648

While the introduction of explicit sensitivity analysis for each model is an improvement on PR19, the design of this exercise is, at present, a major missed opportunity.

As set out in our submission in January 2023, a better alternative approach is to focus sensitivity tests on the sensitivity of the *predicted values* calculated from a model's estimated coefficients: i.e. on the sensitivity of "modelled costs". There are firm grounds for doing so:

- **Relevance**. It is much more relevant to the way that the econometric models are to be used. For the purposes of Ofwat's price control determinations, the key results from econometric models that are actually used are the modelled costs (which are calculated in light of each estimated coefficient) rather than individual coefficients taken in isolation.

- **Practicality**. It provides a way to condense information on sensitivity in a logical way. Measures of the precision or sensitivity of modelled costs will reflect the precision or sensitivity of individual coefficients, with coefficients weighted according to their relative influence on modelled costs.

The CMA has also recognised the logic for looking at information on the precision or sensitivity of modelled costs, rather than simply looking at results for individual coefficients. In the Bristol Water redetermination in 2015, the CMA drew on evidence relating to the confidence intervals for predicted costs:[8]

> *"We considered it important to keep in mind that the purpose of the econometric model was as a means to estimate a water company's base expenditure requirements. Rather than focusing on the estimated variance for specific coefficients, we also calculated the variance (or confidence interval) for the estimate of expenditure derived from these models. This reflects the combined effect of the variance of each of the estimated coefficients in a model on the prediction derived from them."*

While we follow the CMA in giving attention to modelled costs, we depart from the approach used by the CMA in 2015 by focusing on the *observed sensitivity* of modelled costs to minor dataset variations, rather than on statistical estimates of the variance or confidence intervals around modelled costs. We consider that this departure has major benefits. It recognises the logic, especially with relatively small samples of non-independent observations, of looking at the sensitivity of key model estimation results to minor variations in the dataset – an approach which has a foundation in statistical literature,[9] as well as considerable intuitive appeal. Furthermore, this approach avoids the use of statistical concepts such as confidence intervals that rely on assumptions that do not necessarily apply in practice.[10]

In the table below, we compare the sensitivity tests used by Ofwat so far in the PR19 process with a metric for modelled cost sensitivity developed by Reckon (we referred to this in our model submission to Ofwat in January 2023). A more detailed description of the calculation of that metric is contained in Appendix 1 to Reckon's report for Bristol Water and Wessex Water (2023) *Residential retail cost assessment at PR24: econometric benchmarking models.*[11]

---

[8] (2015) *Bristol Water plc: A reference under section 12(3)(a) of the Water Industry Act 1991 Appendices 1.1 – 4.3*, paragraph 215.
[9] This type of approach relates closely to the long-established technique of Jackknife resampling in the statistics literature.
[10] For example, assumptions on a large same size, the shape of probability distributions or the assumptions used to transform confidence intervals around predicted values for the dependent variable into predicted value for modelled costs in £m or £ per customer.
[11] https://www.reckon.co.uk/s/2023-04-06-Reckon-Benchmarking-models.pdf.

Wessex Water
Claverton Down
Bath BA2 7WW
Tel 01225 526 000
Web wessexwater.co.uk

Table 1 Comparison of Ofwat sensitivity tests to Reckon metric of modelled cost sensitivity

Attribute of metrics	Ofwat sensitivity tests from April 2023 consultation	Reckon metric for modelled cost sensitivity
Provides information on the sensitivity of model estimation results to changes in the dataset by dropping companies and/or years	Yes, but information on sensitivity is limited by tests looking only at the impact on results from dropping the first and last year of data and the most efficient and least efficient company	Yes. This metric is calculated using estimation results from hundreds of versions of the original dataset which allow for the effects of dropping different combinations of each company and each year from the dataset
Attention to those model estimation results that are most important for the purpose for which the econometric models are used	Not focused on the sensitivity of modelled costs The sensitivity tests do not capture changes in the scale/magnitude of estimated coefficients other than changes that lead to a change in the sign of the coefficient. The sensitivity tests give the same weight to the effects on coefficients for explanatory variables that have relatively little impact on model estimation results as those which have a large impact. The sensitivity tests give weight not just to the sensitivity of coefficients but to the sensitivity across thresholds for statistical significance, which is of questionable importance	Yes, this metric is focused on the sensitivity of modelled costs The modelled costs are the key results that are taken from the econometric models and used for the purposes of setting price control allowance. The sensitivity of modelled costs will reflect the sensitivity of individual coefficients, weighted by their relative importance in the determination of modelled costs.
Usefulness for the practical purposes of model assessment and selection	Ofwat's sensitivity tests lead to a model being rated red / amber / green for four individual dataset variations. This limits the usefulness for model assessment and selection purposes. For example, the majority of Ofwat's consultation models are rated "G" for all of the sensitivity tests, providing no practical basis to understand differences in the sensitivity between these models.	A single metric for each model that can be used to compare that model to alternative models with greater granularity and informational power than a red/amber/green categorisation
Compromised by perverse results in some cases	Yes. For example, under Ofwat's sensitivity tests, as explained in its guidance from November 2022, a model might be given an amber rating when the statistical significance of a coefficient falls from 1% to 5%, or increases from 5% to 1%, even if the estimated coefficient value stays the same.	We are not aware of perverse results from the metric
Practicality	This test is not available via off-the-shelf software such as Stata. We have direct experience in implementing Ofwat's sensitivity tests from the process ahead of the model submissions in January 2023. We found these tests to be by far the most difficult and time-consuming to apply across the set of tests requested by Ofwat. The test seemed particularly difficult to automate via Stata code. Ofwat did not provide code that provides for automation of the test.	This metric is not available via off-the-shelf software such as Stata Some bespoke code is needed to calculate the metric However, Reckon can provide Stata code and royalty-free licence to Ofwat to implement the metric (subject to acknowledgement in code that it was developed by Reckon)

Wessex Water
Claverton Down
Bath BA2 7WW
Tel 01225 526 000
Web wessexwater.co.uk

Wessex Water Services Limited registered office
Registered in England No 2366648

Our contention is not that the Reckon metric for modelled cost sensitivity is necessarily the best possible metric; there may be scope for further refinement. But it is far better than the sensitivity tests used by Ofwat so far and, for the purposes of Ofwat's assessment of model robustness, and as it further develops and selects its econometric models for PR24, it is of high importance relative to most of the other tests and metrics currently reported by Ofwat.

A more usable and useful metric of goodness of fit

In addition to the sensitivity of modelled costs, another key aspect of Ofwat's concept of model robustness is goodness of fit.

Ofwat rightly recognises the importance of goodness of fit in its approach. It has two different metrics that can be used – in some circumstances at least – to compare the goodness of fit of alternative models: R-squared and the range of efficiency scores. However, as Ofwat recognises, there are limitations with each of these.

In the table below, we compare the two measures of goodness of fit used within Ofwat's current approach against an alternative normalised metric of goodness of fit developed by Reckon (we referred briefly to this in our model submission to Ofwat in January 2023). A more detailed description of the calculation of this alternative metric is contained in Appendix 1 to Reckon's report for Bristol Water and Wessex Water (2023) *Residential retail cost assessment at PR24: econometric benchmarking models.*[12]

Table 2 Comparison of alternative metrics of goodness

Attribute of metrics	R-squared	Range of efficiency scores	Reckon normalised RMSE metric
Can be used to compare goodness of fit for models with different dependent variable?	No	Yes	Yes
Can be used to compare goodness of fit between results derived from suites of multiple models?	No – defined only for individual econometric models	Yes	Yes
Takes account of goodness of fit of full data sample?	Yes	No, focused only on extreme values	Yes
Grounded on established statistical concepts and approaches for goodness of fit?	Yes	The spread between max and min residuals is a possible measure of goodness of fit but not a standard measure	Yes, is a version of RMSE tailored to concept of fit that is most relevant (outturn spend relative to modelled costs)

[12] https://www.reckon.co.uk/s/2023-04-06-Reckon-Benchmarking-models.pdf.

Attribute of metrics	R-squared	Range of efficiency scores	Reckon normalised RMSE metric
Practicality	Directly provided as part of standard estimation results	Not provided as part of standard estimation results and requires a set of further calculations.	Not provided as part of standard estimation results Reasonably simple calculation once modelled costs across companies and years have been derived from regression Reckon can provide Stata code and royalty-free licence to Ofwat to implement the metric (subject to acknowledgement in code that it was developed by Reckon)

Examples of practical benefits

We highlight below several specific insights that we have drawn through use of these metrics in our own review of Ofwat's consultation models (these are limited by the time available for our response):

- These metrics provide evidence of the inferior performance of the new MSOA-based density metric which seems to be missed from Ofwat's sensitivity tests and review of coefficient p-values.
- These metrics reveal what appears to be a problem with the TWD5 model in terms of modelled cost sensitivity, which is greater for this model than for Ofwat's other treated water distribution models.
- These metrics show that, in the context of its sewage treatment models, Ofwat's new weighted average treatment complexity variable seems to perform better than the two treatment works size variables, in terms of both goodness of fit and modelled cost sensitivity.
- These metrics also provide evidence of how adding time trends seems to improve performance of each of Ofwat's sewage treatment models, in terms of both goodness of fit and modelled cost sensitivity.
- These metrics show that, within the context of Ofwat's bioresources models, its proposed unit cost approach seems to consistently reduce modelled cost sensitivity, without any significant effect on goodness of fit, compared to the corresponding aggregate cost models.

We elaborate on the points above in the relevant parts of Appendix A2 and A3.

A2: Changes in costs over time and time trend variables

Ofwat's consultation presents results for over 60 different econometric benchmarking models. The wholesale models are estimated using a panel data set spanning 11 years, while the residential retail models draw on data spanning 9 years. There is considerable evidence that, over this timeframe, industry-wide costs (relative to CPIH) have changed considerably, in ways that are not explained by movements in the cost driver variables captured in Ofwat's econometric models.

None of Ofwat's consultation models have a time trend. Ofwat does not explain the exclusion from consideration of any models with time trends by reference to evidence on the statistical performance of such models. Instead, it rejects time trend variables as a matter of principle.

Our view is that Ofwat's approach and position on the use of time trends in its econometric benchmarking models are not sustainable. It is not evidence-based, and it conflicts with standard statistical/econometric practice for time series panel data.

Our position is not that all parts of the value chain warrant time trend variables, but that this should not be dismissed as a matter of principle and should instead be informed by case-specific evidence from the estimation of models with time trends.

Appendix A2 is organised as follows:

- Statistical support for time trend variables.
- Omission of time trend compromising the modelling of cost driver variables.
- Engineering, operational and economic rationale for a time trend.
- Setting allowances for the 2025-30 period in the light of models with a time trend.
- Ofwat's reasoning for rejecting the use of time trends.

Statistical support for time trend variables

We have not carried out a detailed review across all parts of the value chain, across all the different model specifications used by Ofwat, but have found some examples of where there is clear support for time trend variables being incorporated within Ofwat's current model specifications.

One example is residential retail costs. For both the models of bad debt related costs and the models of other retail costs, we found that the time trend variable was significant at 1% or 5% confidence levels for our preferred model specifications. We also found that the coefficients and statistical significance of other candidate explanatory variables were distorted by the omission of a time trend. We provide further information on the impacts of time trends on the modelling of residential retail costs in the subsection further below.

Another example is the set of three sewage treatment models from Ofwat's consultation. We examined the performance of versions of these models which included a time trend. Our key findings in terms of conventional results for these are as follows (see the table below for more detailed model estimation results):

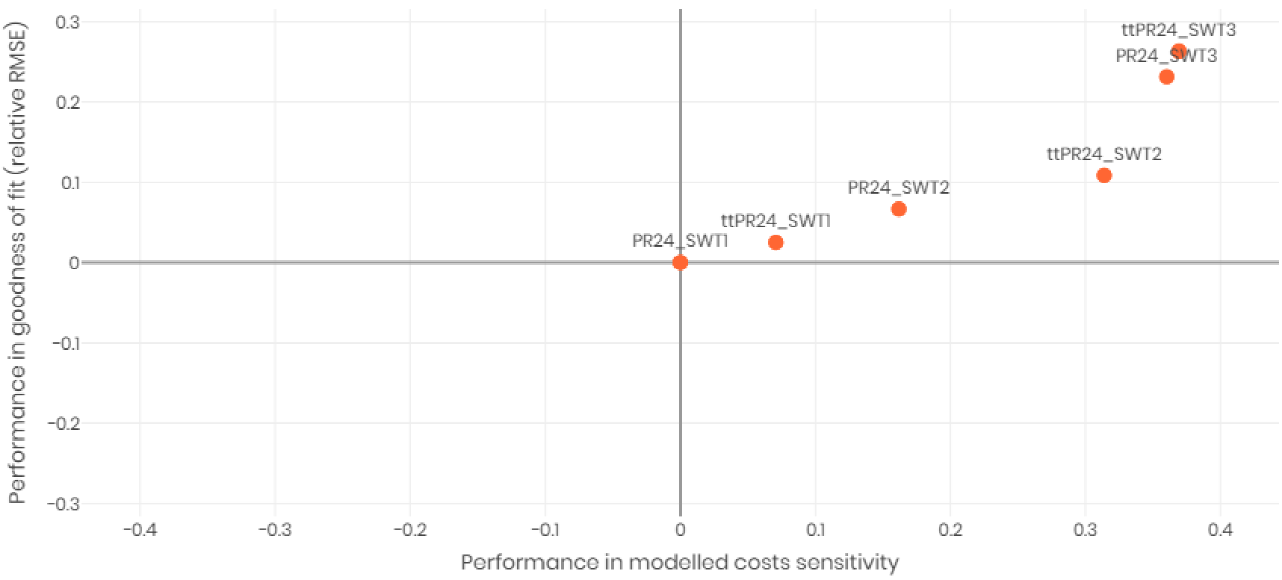
- The time trend variable is statistically significant at the 10% level, with p-values of around 6% to 7%.
- The magnitude of the time trend is quite consistent across the three model variants.

We also compared the time trend versions using the enhanced metrics of modelled cost sensitivity and goodness of fit introduced earlier in this response. We summarise the results from this comparison in the chart below. To help explain this chart (we use some similar charts in Appendix A3).

- The chart covers six sewage treatment models, comprising Ofwat's three consultation models and three variants on these with a time trend added (and no other changes). The model denoted with the "tt" prefix is the version of the corresponding Ofwat model with a time trend added.
- The charts show the relative performance of models in terms of the Reckon metrics of (a) modelled cost sensitivity (on the x axis) and (b) normalised goodness of fit (RMSE) on the y axis.
- This chart takes model SWT1 as a reference point, with the modelled cost sensitivity and goodness of fit for the other five models calculated relative to the corresponding figures for SWT1.
- Models that feature closer to the right-hand side of the chart are better in terms of modelled cost sensitivity and models that are closer to the top of the chart perform better in terms of goodness of fit.

We can see that, for each of Ofwat's three sewage treatment models, the corresponding version of it with a time trend is found, on these metrics, to perform better in terms of both modelled cost sensitivity and goodness of fit.

Figure 1 Comparison of goodness of fit and modelled cost sensitivity for SWT models with/without time trend



Wessex Water
Claverton Down
Bath BA2 7WW
Tel 01225 526 000
Web wessexwater.co.uk

Wessex Water Services Limited registered office
Registered in England No 2366648

Table 3 Estimation results for Ofwat's SWT models and versions of these with time trends

Model Ref	SWT1	SWT2	SWT3	ttSWT1	ttSWT2	ttSWT3
Dependent variable	Ln(Sewage treatment botex plus)					
Explanatory variables						
Load (log)	0.653*** (0.000)	0.723*** (0.000)	0.788*** (0.000)	0.728*** (0.000)	0.808*** (0.000)	0.839*** (0.000)
Load treated with ammonia consent ≤ 3mg/l	0.006*** (0.000)	0.006*** (0.000)	0.006*** (0.000)	0.004*** (0.003)	0.004*** (0.003)	0.004*** (0.001)
Load treated in size bands 1 to 3 (%)	0.029 (0.211)			0.039 (0.189)		
Load treated in STWs ≥ 100,000 people (%)		0.008*** (0.007)			-0.010*** (0.001)	
Weighted average treatment size (log)			-0.242*** (0.000			-0.238*** (0.000)
Time trend				0.012* (0.068)	0.013* (0.063)	0.012* (0.057)
Overall R-squared	0.85	0.87	0.91	0.87	0.88	0.92
Observations	110	110	110	110	110	110

We also briefly reviewed the impact of adding time trend variables to Ofwat's six consultation models for treated water distribution. We found as follows:

- Across all six models, the time trend variable is significant at least at the 5% level.
- The inclusion of the time trend variable did not seem to adversely affect the statistical significance of the coefficients on the cost driver explanatory variables.

Omission of time trend compromising the modelling of cost driver variables

In some cases, a time trend variable can play an important role as a "control variable" within an econometric model which, if omitted, can distort other model estimation results.

In the case of Ofwat's econometric benchmarking models, there is a serious risk that the modelling of cost driver relationships is being compromised:

- **Loss of valid cost driver variables**. A candidate cost driver variable might be found not to be statistically significant in a model that omits a time trend. This could lead to the loss of that variable from the benchmarking analysis. However, the variable might be statistically significant in a model with a time trend. If Ofwat excludes time trend models as a matter of principle, then this can lead to the loss of quite valid cost driver variables.
- **Inclusion of questionable cost driver variables**. A candidate cost driver variable might be found to be statistically significant in a model that omits a time trend, but not statistically significant once a time trend is included. If Ofwat excludes consideration of time trend models as a matter of principle, then it

might include this variable in its models even though there is a questionable basis for the variable. For instance, there may be no significant underlying cost causality relationship for the variable, but it might simply be a proxy for – or make up for – the omission of the time trend.

It seems that the first of these problems has arisen, in practice, in the case of Ofwat's development of residential retail models. We elaborate on this below. (We also think that there is evidence of the second problem arising in the case of Ofwat's residential retail models, in relation to the specific metrics used to capture deprivation and arrears risk and we cover this in Appendix A3 below in response to Ofwat's consultation questions its retail models.)

At PR19, Ofwat's models of other residential retail costs included an explanatory variable for meter penetration, which was significant at the 1% level. This variable made intuitive sense and improved incentives.[13] A striking feature of Ofwat's consultation models for residential retail costs at PR24 is that none of them include the meter penetration variable. This seems to be a direct consequence of Ofwat's models not controlling for changes over time in industry-wide costs.

We can see this in the set of model estimation results in the tale below. Models ROC1 and ROC2 reproduce the models from Ofwat's consultation. We can see that if we simply add meter penetration to these models (i.e. models ROC1mp and ROC2mp in the table), the meter penetration variable is not statistically significant at the 10% threshold. However, if we add both a time trend and a meter penetration variable to Ofwat's models (i.e. models ROC1mptt and ROC2mptt) we find that the meter penetration variable is significant at either 1% or 5% (depending on the model) which would be strong grounds to support it given Ofwat's approach. Furthermore, the time trend variable is significant at the 5% level.

Table 4 Estimation results for Ofwat models of other retail costs and some variants on these

Model Ref	ROC1	ROC2	ROC1mp	ROC2mp	ROC1mptt	ROC2mptt
Model type	Ofwat consultation model		Ofwat + meter penetration		Ofwat + meter penetration and time trend	
Dependent variable	Ln (Other retail costs per household, CPIH adjusted)					
Explanatory variables						
Proportion of dual service customers	0.002** (0.029)	0.003*** (0.001)	0.002** (0.025)	0.003*** (0.000)	0.002** (0.023)	0.003*** (0.002)
Proportion of metered connections			0 (0.809)	0 (0.834)	0.007** (0.012)	0.007*** (0.008)
Ln (Total households connected)		-0.045 (0.139)		-0.049 (0.117)		-0.039 (0.280)
Time trend					-0.022** (0.014)	-0.022** (0.013)
Overall R-squared	0.12	0.13	0.13	0.14	0.14	0.15
Observations	153	153	153	153	153	153

[13] Given that there are some incremental costs of serving metered customers (e.g. meter reading and additional customer queries) a regulatory approach which provides zero incremental revenue per customer who becomes metered does not offer good incentives for companies to increase meter penetration.

This example shows how the omission of a time trend variable can lead to other variables being estimated more imprecisely and to the selection of models that have worse coverage of the underlying cost drivers.

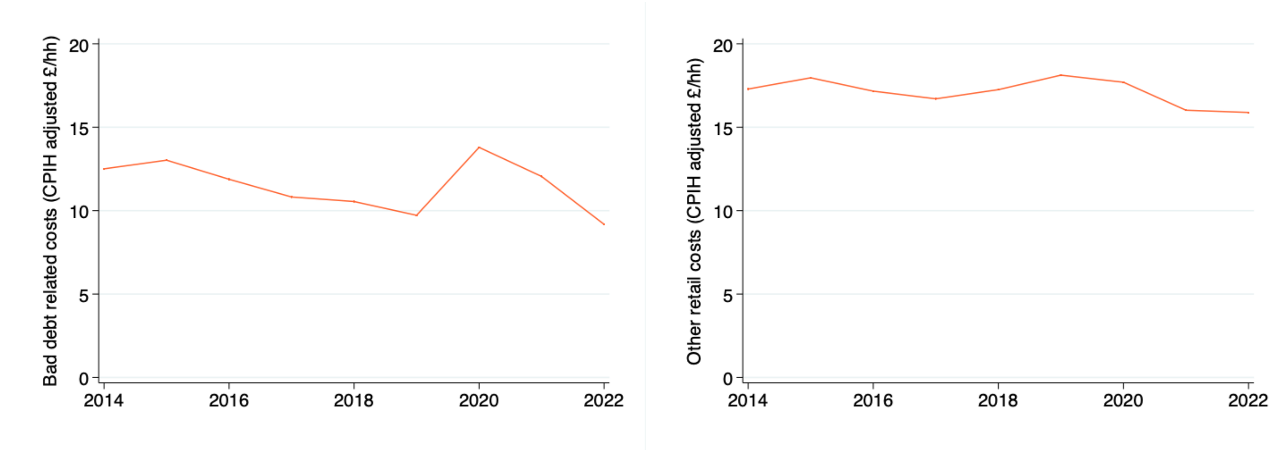
Engineering, operational and economic rationale for a time trend

When applying a time trend it is relevant to consider whether there are economic, operational or engineering factors that could explain a material non-zero time trend relative to CPIH. Our view is that this is clearly the case.

In the case of residential retail costs, we found that time trend variables were consistently statistically significant and implied reductions in costs over time, relative to CPIH. A key rationale that we see for this is the idea that, as a consequence of Ofwat's PR14 price control reforms for residential retail activities, there have been relatively high levels of efficiency improvement over time in residential retail across the industry. While we do not believe that that scale of efficiency improvement will be sustainable in the long term, it is something that is highly relevant to capture in the modelling applied to historical data.

The chart below shows the profile over time for the two different categories of residential retail costs used in Ofwat's modelling. The chart on the left is for bad debt related costs and the right is for other retail costs. The charts show total costs across all companies in the industry, expressed on a per customer basis. We see a downward trend in both charts (albeit with Covid-related effects in 2019/20 and 2021/22 for bad debt related costs).

Figure 2: Profile over time of total industry bad debt related costs and other retail costs per household



In the case of wholesale water and wastewater network plus activities, where we found that time trend variables were statistically significant these implied increases in costs over time, relative to CPIH. We do not think that there is one unique factor that is likely to explain these changes over time in industry-wide costs (beyond factors captured in the cost models). There are likely to be multiple factors at play (some of which might go in opposite directions). Some key examples include the following:

- **Increases to industry-wide costs from actions to improve performance in response to the incentives arising from PCs with financial ODIs**. For example, companies may incur higher operational costs each year to sustain improvements in water supply interruptions (indeed this is Wessex Water's own experience). While Ofwat does not seem to have given much weight in the past to the idea that performance improvements may go hand-in-hand with increases in efficient levels of costs, it is important to recognise that such a scenario is an intended effect of its regulatory incentive model implemented via PCs and ODIs: these are intended to encourage companies to incur additional costs to improve customer service and environmental performance provided that the increases in costs do not exceed the estimated benefits to customers of those improvements.

- **Increases to ongoing operational and maintenance costs as a consequence of customer and environmental improvements achieved via historical enhancement schemes approved and funded by Ofwat (size of asset base**). For example, there may be opex and capital maintenance arising from past enhancement capex driven by WINEP that is not captured by model explanatory variables or ODIs / PCs.

- **The net effect of industry-wide productivity growth and RPEs differing from 0% per year**. Both Ofwat and the CMA set wholesale allowances at PR19 under an assumption of base costs falling relative to CPIH (once allowance is taken for cost drivers including population growth). Ofwat's own decision-making involves the view that there are reasons for efficient cost trends not to match CPIH over time.

It is not possible for time trend variables to differentiate between the effects above but other sources of evidence might be informative in trying to understand the likely balance between them in understanding the net changes over time.

In its report for Ofwat, CEPA said that a time trend "does not address the root cause behind increasing expenditure" and suggested the alternative of "Controlling for additional cost drivers that explain the increase in expenditure more precisely (e.g., including a leakage variable)".[14] We think that, in practice, it will be difficult to try to properly capture the types of effects listed above through the inclusion of additional cost drivers in the econometric models. We would welcome the chance to review any models proposed by Ofwat or CEPA that are intended to achieve this. In the absence of such models, it seems unreasonable to refer to this alternative approach as a reason against using time trend variables.

Setting allowances for the 2025-30 period in the light of models with a time trend

We recognise that some of the concern with the use of time trend variables in the models applied to historical data is the potential implications of this for the setting of allowances for the 2025-30 period.

A key point to emphasise is that inclusion of the time trend in a model estimated on historical data does not require that the price control allowances set for the 2025-30 period are calculated in a way that simply extends the estimated time trend into that period (a concern that Ofwat seems to have).

Even if we think that a time trend should be included in an econometric model that covers a historical data period, on empirical grounds, this does not mean that we would necessarily look to apply that time trend in making projections for the future. We would want to understand more about the likely drivers of the observed time trend in the past period and their likely relevance to the future period. There may in some cases be good reasons for thinking that the rate of change in costs observed historically is unlikely to be experienced in the forthcoming five-year price control period.

It is perfectly possible to use a time trend to model historical costs and then to adjust or "disapply" that time trend when projecting into the future. See the report that we commissioned from Reckon on *Residential retail cost assessment at PR24: projection of cost benchmarks* (April 2023)[15] for a detailed practical

[14] CEPA (2023) *PR24 Wholesale Base Cost Modelling*.

[15] https://www.reckon.co.uk/s/2023-04-06-Reckon-Benchmark-projection.pdf.

explanation of how projections can be made for a future price control period in a way that does not assume continuation of the time trend estimated via econometric models applied to past data.

Nonetheless, the time trend estimated on historical data provides valuable information on the evolution of industry-level costs over time (due to factors not captured by cost driver explanatory variables). This information is highly relevant to setting price controls for the 2025-30 period.

In contrast, Ofwat's current approach to econometric modelling – across wholesale and retail activities – has the effect of burying evidence on the changes over time in costs that have been observed.

Ofwat's reasoning for rejecting the use of time trends

Ofwat's consultation does not provide any detailed consideration of the case for including time trend variables in some of its models. However, Ofwat did respond to suggestions from Wessex Water and South West Water for the inclusion of time trend variables.

We welcome Ofwat's engagement with this issue, which has had relatively limited attention within the industry. We provide a response below to the points raised by Ofwat against the use of time trends.[16] Overall, this demonstrates that there is no valid basis for the objections to time trends raised by Ofwat.

Table 5 Response to Ofwat argument against the use of time trend variables

Element of Ofwat argument	Response
We have not seen clear evidence to explain the increase in expenditure [implied by time trend variables], and there is a risk that a time trend captures factors that are inside of company control	There is no basis for the idea – adopted by Ofwat here – that for an explanatory variable to be included in the econometric benchmarking models there must be clear external evidence (i.e. outside of the model estimation results) to support that variable's inclusion. This test is not applied to other variables that are considered in the modelling. For other variables (e.g. the new variables for coastal population and urban rainfall) it seems sufficient that there is *intuitive rationale for considering the variable* – after that it is an empirical matter in light of the results from models including the variable. We consider that there is a *clear intuitive rationale* – reflecting engineering, operational and economic considerations – *for considering a time trend variable* – after which it is an empirical matter in light of the results from models including the time trend variable. See the subsection on this above. More generally, it is not standard practice in econometric modelling on time series data or time series panel data to only include a time trend if there is "clear evidence" from outside the model estimation results to support the inclusion of that variable. Instead, it is for the econometric modelling exercise to help inform on whether the time trend variable should be used. Ofwat cites a "risk that a time trend captures factors that are inside of company control" but gives no examples of what these factors might be and why they would render a time trend inappropriate. Given the econometric models are estimated across a number of independent companies, a statistically significant time trend coefficient is unlikely to arise from the decisions of one or two companies and is more likely to reflect the effects of industry-wide phenomena.

[16] This is based on Ofwat's comments on page 31 of its consultation in the subsection "Time trend, year dummies and other dynamic factors". Ofwat does make some further references to time trends in the consultation but these do not go beyond what is covered in this subsection.

Element of Ofwat argument	Response
There is also a risk that the increase in wholesale water base expenditure in recent years is not permanent or will continue at the same rate	We have some sympathy with the concern raised here, but it is misplaced. The inclusion of a time trend in the modelling of historical costs does not force Ofwat to apply that time trend when setting allowances for future periods. If Ofwat thinks that any increase in costs observed historically will not be permanent, or will continue at the same rate, then it can take account of this in the way it sets allowances for the forthcoming price control period. Such expectations about the future do not provide a valid basis for rejecting the use of a time trend in modelling historical expenditure. For further explanation, see the subsection above "Setting allowances for the 2025-30 period in the light of models with a time trend". Ofwat's established approach is already "flexible" in the sense that the rates of change over time in costs it assumes for the forthcoming price control period do not match the rates of change in costs that it (implicitly) assumes in the specification of its econometric models. Ofwat's econometric models for PR19 involve the implicit assumption that companies' efficient costs are constant over time relative to CPIH (in the absence of changes driven by cost driver variables). However, when it set wholesale base cost allowances at PR19, Ofwat adopted a different assumption – that wholesale base costs change over time at a different rate to CPIH+0% due to the combined effects of ongoing productivity and RPEs (which had an expected net impact of approximately CPIH-0.5% at PR19 given Ofwat's assumptions).
We prefer to focus on cost drivers that are exogenous	The variable that is used to implement a time trend in the types of econometric models used by Ofwat (the "fye" variable in the dataset published by Ofwat) is clearly exogenous. Unlike many of the variables in Ofwat's consultation models, there is absolutely no degree of company control or influence over the time trend variable.
We prefer to focus on cost drivers that … have a clear engineering, operational, and economic rationale, as set out in our principles of base cost assessment	We believe that there are clear engineering, operational, and economic rationales for considering models with a time trend (as we highlighted elsewhere in Appendix A2). Whether those effects (taken together) have a significant and consistent impact on industry-wide costs is an empirical matter which is best addressed by a review of estimation results from models that include time trend variables, rather than by taking a position – in the absence of evidence – that it is appropriate to use models that assume that efficient costs grow at CPIH+0% (beyond changes driven by cost driver variables). The use of time trend variables is a very common and well-established technique in statistical and econometric practice, and it does not seem valid to reject it on the basis that the time trend does not directly capture the underlying factor(s) that have given rise to the changes observed over time.

A3: Our response to the consultation questions

Q3.1: Do you agree with our proposed set of wholesale water base cost models?

Strongly disagree.

There are a number of elements of the wholesale water base cost models that we support. However, we do not consider the set of consultation models to be suitable for PR24 as they stand.

Ofwat supports its models by reference to the procedures used to develop and test them. However, as discussed in detail in Appendix A1, we see significant limitations in Ofwat's approach to model robustness assessment. This seems to (a) give too much emphasis to pass/fail tests that are not particularly useful for the purposes of assessing the relative performance of alternative models under consideration and (b) miss out on good ways to capture the most important aspects of model robustness and model performance. Improvements to this aspect of Ofwat's approach will be important as Ofwat reflects on stakeholder responses to this consultation and if it looks to narrow down its set of models.

As discussed in Appendix A2, we disagree with Ofwat's rejection of time trends as a matter of principle. There seems to be evidence for time trends in some cases, including for treated water distribution.

We make comments below about density variables in response to the separate question on this. But we highlight here that, drawing on analysis across the different model variants that Ofwat is consulting on, we found that for water resources plus the models with the "Weighted average density – MSOA" variable performed considerably *worse* statistically than those with the "Weighted average density – LAD from MSOA" metric, especially in terms of the sensitivity of modelled costs to dataset variations.

The range of efficiency scores implied by Ofwat's wholesale models are far wider than for wastewater network plus. The efficiency scores for water resources plus are particularly widely spread, raising serious questions about the plausibility of these models as any sort of guide to the efficient levels of costs within the industry.

We consider that it will be important to use a diverse set of models at PR24, reflecting a range of alternative explanatory variables.

We are concerned about the inclusion of growth-related expenditure without further information from Ofwat as to how it will (a) adjust for differences between companies and over time in the drivers of such expenditure; and (b) and how it will deal with potential lumpiness of such spend between price control periods in a context where its practice of asymmetric cost sharing rates prevents over-spends and under-spends against modelled allowances cancelling out over time.

Ofwat's approach towards the treatment of enhancement operating expenditure, while somewhat improved, is still not reasonable, with too much expenditure prone to being wrongly excluded from the base modelling. Ofwat said in its final methodology that it would include enhancement operating expenditure where it was "more certain the costs are ongoing" given concerns about double counting. But only a narrow set of enhancement operating expenditure has been included. We urge Ofwat to issue a targeted information request to companies so that the element of enhancement operating expenditure that is "ongoing" can be informed by evidence rather than guesswork. This is something that Ofwat should do for its PR24 modelling – we see no justification for Ofwat's plan to "look to improve enhancement opex reporting further ahead of the 2025-26 reporting year". See section 5.4 of the report we commissioned from Reckon in 2022 for practical suggestions on improved reporting for enhancement operating expenditure.[17]

Q3.2: Do you agree with the inclusion of average pumping head in a sub-set of treated water distribution and wholesale water models?

Neither agree nor disagree – N/A.

Ofwat has said that it is still concerned about the quality of APH data.[18]

Likewise, we continue to have reservations about the quality and consistency of data on average pumping head**.** The inherent complexity of the APH variable raises some data quality risks, with potential for inconsistencies and possible mistakes in its application. The number of booster pumping stations variable is by its nature a much simpler variable that is less open to data quality issues. While some efforts have been made in the industry to improve APH data there is limited ability to change historical APH data, yet this will receive considerable weight in the modelling.

Nonetheless, we agree with Ofwat that there are grounds for the inclusion of the average pumping head variable in a *subset* of the models for treated water distribution and wholesale water models, alongside models that include the variable used by Ofwat and the CMA to proxy for topography at PR19: the number of booster pumping stations per length of mains.

In relation to comparisons between APH and variables relating to the number of booster pumping stations, we note that the latter has the potential to capture more than just topography – it may also pick up on density effects which are not fully captured already under the relevant density metrics in the model (as more rural areas may have more booster pumping stations).

In our submission in January 2023, we found average capacity of booster pumping stations to be a good driver of TWD costs and was included in one of our submitted TWD models. Using the enhanced metrics discussed in Appendix A1, we found that this could offer slightly reduced modelled cost sensitivity with no significant impact on goodness of fit (e.g. using this to replace the number of booster pumping stations in in models TWD1 and TWD2).

We also highlight that Ofwat's model TWD5 (which includes the average pumping head variable) seems to perform considerably worse than its other treated water distribution models. In particular, we found that the modelled costs for TWD5 were more than twice as sensitive to variations in the dataset (dropping years and companies on a systematic basis) as the other treated water distribution models. We caution against TWD5.

Q3.3: Do you agree with our approach to modelling population density? Which of the three explanatory variables do you support? (a) Weighted average density – LAD from MSOA; (b) Weighted average density – MSOA; (c) Properties per length of mains.

Neither agree nor disagree – N/A.

First and forecast, we see density as a highly important cost driver to capture in the econometric models of wholesale base-plus expenditure. We are pleased to see that all of Ofwat's wholesale water models incorporate at least one explanatory variable for density and make use of linear and squared terms.

[17] https://www.reckon.co.uk/s/2022-09-01-Base-enhancements-report.pdf

[18] Ofwat (2023) *Econometric base cost models for PR24*, page 23.

Ofwat's modelling of the density cost driver improved at PR19, drawing on contributions from Wessex Water concerning the introduction of variables based on weighted-average population density metrics and the use of both linear and squared terms.

All else equal we favour exogenous density variables. Nonetheless, we see value in considering endogenous variables for density (e.g. properties per length of mains) and potentially including them, alongside exogenous variables, in the overall model suite if they perform considerably better than the exogenous variables in terms of statistical performance.

For the consultation, Ofwat has developed new metrics of density that draw on granular ONS population density.

For the type of weighted average population density variable used in Ofwat's consultation, there are two interrelated ways that the granularity matters:

- **Mapping company boundaries**. What level of granularity is used to map geographical units (for which ONS population data is available) to water company boundaries?
- **Calculating population density**. What level of granularity is population density to be calculated at, before the weighted average is calculated for each water company area?

For PR19, Ofwat used the LAD as the level of granularity for both of these elements.

Our reading of Ofwat's consultation is that the variables that it is consulting on both use the MSOA data for mapping company boundaries but that one involves calculating population density at the MSOA level and the other uses a proxy approach for calculating population density at the LAD level.

A further option, which has been used by Reckon in the past, is to use the LSOA-level data for both the mapping and the calculation of the metric of interest.[19]

Nonetheless, we would disagree with the idea – perhaps suggested in Ofwat's consultation document – that calculating population density at a more granular level before taking a weighted average is necessarily better. While there are benefits from using density metrics that allow for a better mapping of ONS data to water company areas, this is not the only factor that is relevant and may not be the most important. In principle, what matters is the geographic level at which the strongest relationship between population density and efficient costs applies.

In practice, we think that this means the following:

- There is not a good basis for forming an a priori position on the most appropriate density metric(s) to use. Instead, we think that weight should be placed on the findings from a review of the model estimation results from models which differ only in terms of those metrics. In other words, this is best approached as an empirical matter.
- The most appropriate density metric, as informed by the empirical evidence, may differ between different parts of the value chain (e.g. water resources plus versus treated water distribution, versus sewage collection).

[19] See for example Reckon (2018) *Econometric models for residential retail cost assessment*.

- The empirical evidence on the relative performance of alternative density metrics may depend on other aspects of model specification, and it would be prudent to reassess how these metrics perform if other aspects of model specifications change significantly during the PR24 process.

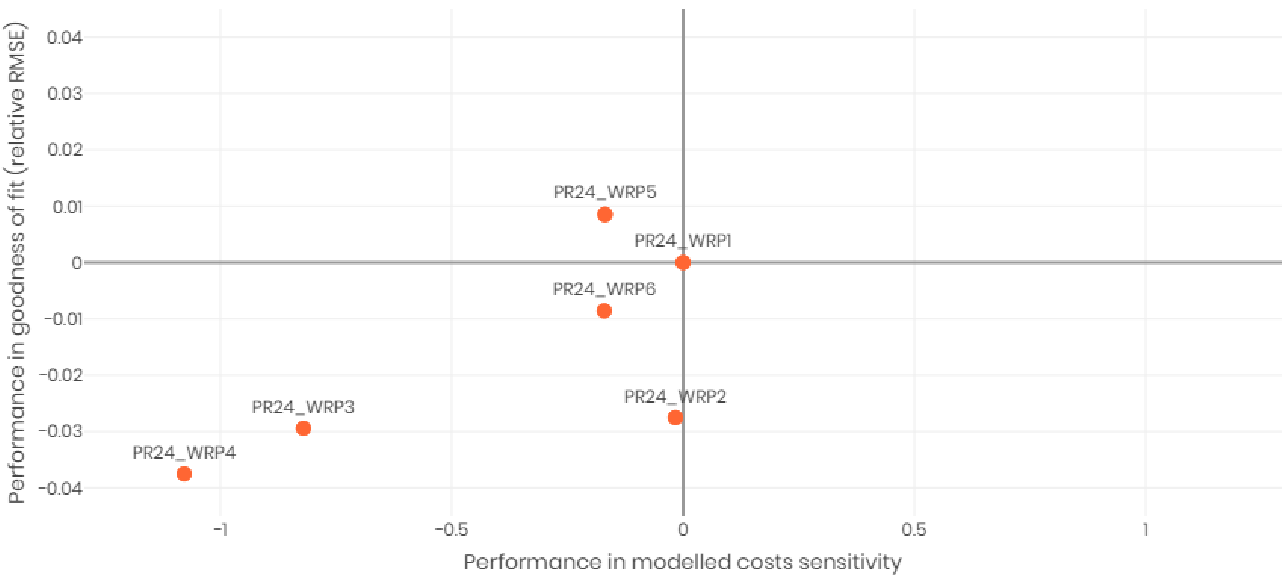
We now present some evidence on the performance of Ofwat's alternative density metrics within the context of its consultation models.

The chart below draws on the metrics of normalised RMSE and modelled cost sensitivity developed by Reckon (see appendix A1) and compares the performance of Ofwat's water resources plus models in these two key dimensions of model performance (taking model WRP1 as a reference point). It shows that models WRP3 and WRP4 (which use the "weighted average density – MSOA" metric) perform considerably worse than the other four WRP models in terms of modelled cost sensitivity, and slightly worse than these models in terms of goodness of fit. For instance, the modelled costs for WRP4 are found to be more than twice as sensitive to minor dataset variations than model WRP1 which indicates a much lower degree of model robustness.

We think that this provides empirical evidence, *in the specific context of Ofwat's models for water resources plus*, to favour the "LAD from MSOA" density metric over the "average density – MSOA" metric.

The more conventional model estimation results, and various tests, reported by Ofwat do not do a good job of revealing the worse performance of models WRP3 and WRP4. We can see that these models have lower R-squared than corresponding models with either the LAD from MSOA" density metric or properties per length of mains metric. But Ofwat's sensitivity tests are not targeted at the most relevant aspects of the econometric results (see Appendix A1 above) and do not pick up on the much greater sensitivity of modelled costs under WRP3 and WRP4.

Figure 3 Comparison of goodness of fit and modelled cost sensitivity for Ofwat WRP models



Wessex Water
Claverton Down
Bath BA2 7WW
Tel 01225 526 000
Web wessexwater.co.uk

Wessex Water Services Limited registered office
Registered in England No 2366648

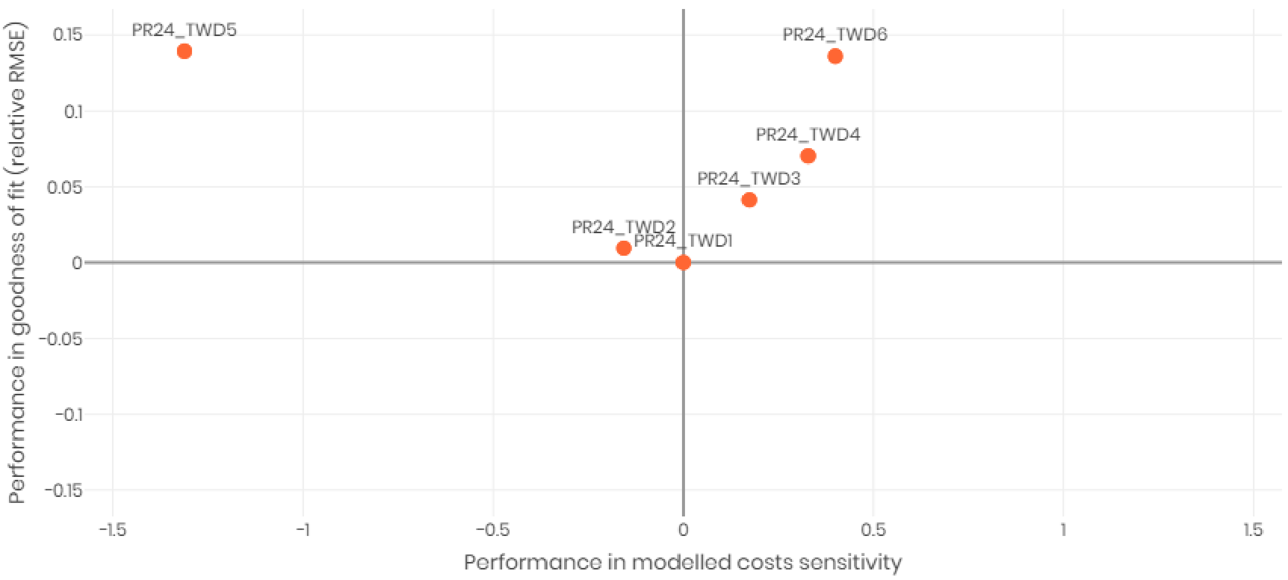
We also considered versions of Ofwat's water resources plus models with its new density metrics replaced by its PR19 LAD-based metric.[20] We carried out some basic analysis and found that the models with the PR19 LAD-based metric performed better in some relevant ways. In particular, the p-value for this density metric seemed to be lower than for Ofwat's new density metrics and the R-squared for models with the PR19 metric was slightly higher. We recognise that Ofwat has concerns with some aspects of the PR19 LAD-based metric but we wonder if it might be possible to produce a revised version, which might outperform both of its MSOA-based metrics. It is possible that for water resources plus, the underlying cost driver relationships are such that it is better to capture population density at a larger level of aggregation than MSOA (e.g. LAD or some alternative to this) yet Ofwat's new way of proxying for LAD via MSOA might have some shortcomings that could be improved upon. But we recognise that this is a complicated area.

Finally, for the water resources plus models, we question the rationale for using the explanatory variable concerning the number of properties divided by water mains length as a density variable. While this variable has some logic in models for treated water distribution it would be a strange outcome for this variable to feature in Ofwat's final set of models for PR24. We suggest a focus on exogenous density variables for water resources plus and, in particular, the LAD from MSOA metric.

We now turn to **treated water distribution**. For the six consultation models, we did not identify strong grounds to favour any one of the density metrics to the exclusion of the others. We did however, find that the new "Weighted average density – MSOA" metric relatively performed badly in the model with the APH variable (TWD5).

The chart below shows the enhanced metrics for modelled cost sensitivity and goodness of fit discussed earlier in this response for Ofwat's six TWD models (taking model TWD1 as a reference point).

Figure 4 Comparison of goodness of fit and modelled cost sensitivity for Ofwat TWD models



[20] This was a simplified and approximate approach (e.g. rolling forward historical values for the LAD metric to more recent years).

Drawing on this analysis, we found as follows:

- Model TWD5 performed relatively badly, relative to the other models, in terms of modelled cost sensitivity. This model has the “Weighted average density – MSOA” variable and the APH variable.
- For models with the booster pumping stations variable to capture topography, the model with the “Weighted average density – MSOA” variable had slightly better goodness of fit and worse modelled cost sensitivity than the model with the “Weighted average density - LAD from MSOA” variable.
- There is greater logic for the number of properties divided by water mains length in Ofwat’s TWD models than its WRP models. We found that, using the enhanced metrics for modelled cost sensitivity and goodness of fit, the two models with the number of properties divided by water mains length (TWD3, TWD6) had better goodness of fit and modelled cost sensitivity than the corresponding models with the exogenous density metrics. However, since water mains length is under some degree of company control, these models are subject to potential data issues and endogeneity concerns (management control) for the density variable, and if used it would be better than these are used alongside models with the more exogenous density metrics.

Q3.4: Do you agree we should collect additional data on the number of reservoirs that are designed as high-risk by the Environment Agency and Natural Resources Wales?

Agree.

As a general matter of principle, we support proportionate efforts to improve the econometric benchmarking models over time.

Furthermore, water resources plus seem to be the part of the value chain where Ofwat’s consultation models perform the least well in explaining differences in costs between companies. For instance, the range of efficiency scores reported by Ofwat for its water resources models is around 2, with efficiency ratios of companies in the barely plausible range of 50% to 200%. This is much worse than for all other areas of Ofwat’s cost modelling. In this context, we support further attention to the way that water resources plus models might be improved, including via collection of additional data.

We do not have any objections to providing data on the number of reservoirs that are designated as high-risk by the Environment Agency and Natural Resources Wales.

However, we think that if there is a material underlying cost driver relating to reservoirs that are designated as high-risk, it would make sense to collect information relating to the capacity of those reservoirs, rather than simply their number. The costs of extra maintenance are likely to be affected by the size of the designated reservoirs. Having data on capacity and number would allow for several different explanatory variables related to high-risk reservoirs to be explored in the development of the econometric models.

We suggest that a sensible approach might be to first collect such data on a best endeavours basis and gauge the likely success of the new data for use within the econometric models before requiring formal reporting of the data on an ongoing basis going forwards. It is possible that the incremental costs arising from designation of reservoirs as high risk is not so great as to be picked up in the econometric models covering water resources plus. In other words, we would want to see some initial evidence that the additional data is potentially useful for econometric modelling purposes before it is required on an ongoing basis and subject to assurance requirements.

Q4.1: Do you agree with our proposed set of wastewater network plus base cost models?

Strongly disagree.

There are a number of elements of the wastewater network plus models that we support including improvements since PR19. However, we do not consider the set of consultation models to be suitable for PR24 as they stand.

Ofwat supports its models by reference to the procedures used to develop and test them. However, as discussed in detail in Appendix A1, we see significant limitations in Ofwat's approach to model robustness assessment. This seems to (a) give too much emphasis to pass/fail tests that are not particularly useful for the purposes of assessing the relative performance of alternative models under consideration and (b) miss out on good ways to capture the most important aspects of model robustness and model performance. Improvements to this aspect of Ofwat's approach will be important as Ofwat reflects on stakeholder responses to this consultation and if it looks to narrow down its set of models.

As discussed in Appendix A2, we disagree with Ofwat's rejection of time trends as a matter of principle. There seems to be evidence for time trends in some cases, with the evidence for sewage treatment especially strong.

In line with our broader support for outcomes-based regulation, we consider that Ofwat's cost benchmarking models should be making as much use as possible of explanatory variables that are truly exogenous rather than variables that are reflective of the inputs used by a company and its past management decisions. Ofwat's consultation models contain a mix of explanatory variables, some of which are truly exogenous and some endogenous (i.e. under some degree of management influence over time). We see a legitimate role for endogenous variables as a fall-back option where the underlying cost drivers are difficult to capture well via exogenous variables. But even in these cases, it makes sense to use models with endogenous variables alongside, rather than in preference to, models with exogenous variables. In this context, we propose that Ofwat expands its set of SWT models to include some with exogenous density variables. Each of its current models for sewage treatment involve variables relating to treatment works size which are not exogenous. In terms of the underlying cost drivers affecting the efficient levels of sewage treatment costs, what matters is the opportunities for economies of scale rather than the actual size of sewage treatment works; exogenous measures of population density are highly relevant to the differences between companies in these opportunities.

In our submission in January 2023, we argued for use of a scale variable based on population equivalent rather than load. The former removes the additional assumption and potential inconsistency around converting pe to load. Using the enhanced metrics discussed in Appendix A1, we found that this could offer slightly reduced modelled cost sensitivity with no significant impact on goodness of fit (e.g. replacing the ln(load) variable with the ln(pe) variable in model SWT3).

We make further comments below in response to the separate question on individual cost drivers (e.g. density, economies of scale at treatment works, treatment complexity, rainfall).

We consider that it will be important to use a diverse set of models at PR24, reflecting a range of alternative explanatory variables.

We are also concerned about the inclusion of some growth-related expenditure as well as drainage and flooding risk enhancement expenditure within the base models, without further information from Ofwat as to how it will (a) adjust for differences between companies and over time in the drivers of such expenditure; and (b) and how it will deal with potential lumpiness of such spend between price control periods in a

Wessex Water
Claverton Down
Bath BA2 7WW
Tel 01225 526 000
Web wessexwater.co.uk

context where its practice of asymmetric cost sharing rates prevents over-spends and under-spends against modelled allowances cancelling out over time.

See also our comments on enhancement operating expenditure under question 3.1.

Q4.2: Do you agree with our approach to modelling economies of scale at sewage treatment works? Which of the three proposed explanatory variables do you support? (a) Percentage of load treated in STWs bands 1 to 3; (b) Percentage of load treated in STWs serving more than 100,000 people; (c) Weighted average sewage treatment works size?

Neither agree nor disagree – N/A.

We briefly summarise our response before expanding on our reasoning:

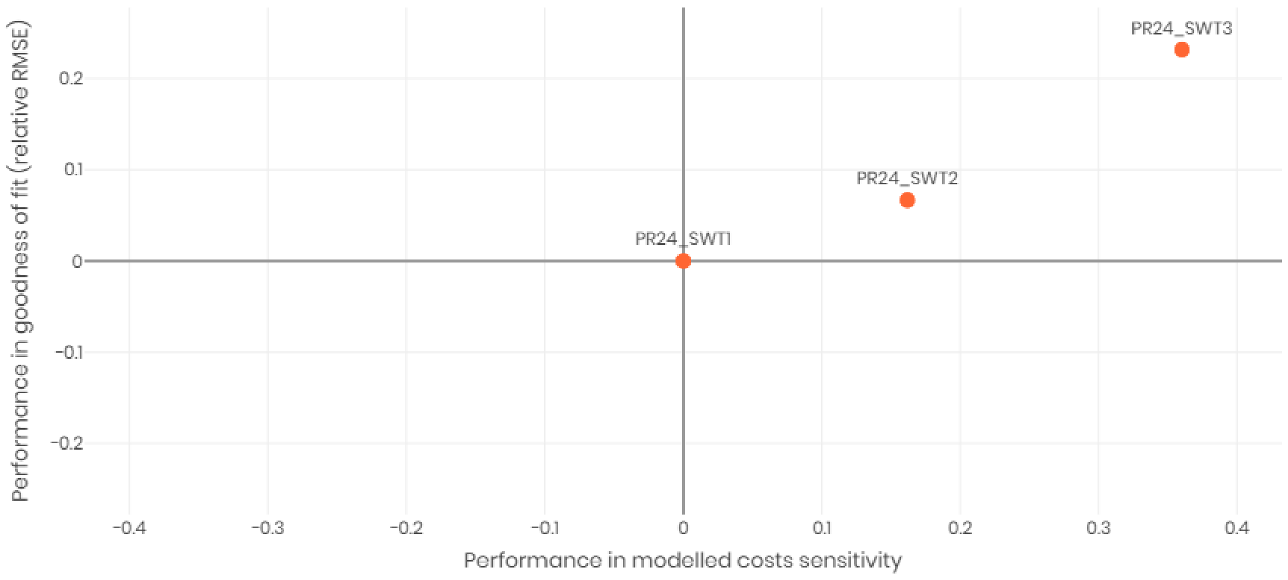
- Across the three approaches that Ofwat consults on, we consider that the weighted average sewage treatment works size variable is preferable, both in terms of its economic and operational rationale, and its statistical performance.
- We propose that, to complement models with the weighted average treatment works size variable, a subset of Ofwat's sewage treatment models use an exogenous population density variable to capture the opportunities that companies face for economies of scale at sewage treatment works. This would fit better with Ofwat's cost assessment principle to focus on exogenous cost drivers.

The weighted average sewage treatment works size (WATS) variable avoids imposing the assumption of a threshold (e.g. treatment works size bands 1-3 vs other band; or works serving more than 100,000 people) above which unit costs are low and below which unit costs are high. We agree with Ofwat that the WATS variable allows for a more gradual relationship between treatment works size and costs.

Furthermore, our analysis indicates that the model with the WATS variable (SWT3) performs better than models SWT1 and SWT2 in terms of both goodness of fit and modelled cost sensitivity (using the enhanced metrics developed by Reckon). We illustrate this in the chart below. This chart takes model SWT1 as a reference point, with the modelled cost sensitivity and goodness of fit for SW2 and SWT3 calculate relative to corresponding figures for SWT1. It shows, for example, that using these metrics, the model with the WATS variable has over 20% better fit with data and more than 30% less sensitivity of modelled costs than the model that uses the variable based on size bands 1 to 3.

Wessex Water
Claverton Down
Bath BA2 7WW
Tel 01225 526 000
Web wessexwater.co.uk

Figure 5 Enhanced model performance metrics for Ofwat sewage treatment models



Turning to the second point summarised above, each of Ofwat's sewage treatment models include an explanatory variable that relates directly to the scale of companies' sewage treatment works. This is an endogenous rather than exogenous variable which is under management control to some degree or in some circumstances.

At present, none of Ofwat's sewage treatment models include exogenous population density variables. This is in contrast to the approach for wastewater collection, bioresources, water resources plus, water treatment and distribution and wholesale water whose consultation models include models with exogenous population density variables,

In its consultation, Ofwat dismisses explanatory variables suggested by other companies on the grounds that these are endogenous and that including them in the models might perversely incentivise companies against engaging in desirable behaviour.[21] We think that Ofwat is right to be alive to such risks, while taking account of what approaches are viable.

We do not advocate a purist approach to Ofwat's cost assessment principle to focus on exogenous cost drivers. There is a need to balance the potential risks from using explanatory variables that are under a degree of management control (e.g. risks of perverse incentives or of overlooking efficient practices) against the potential limitations to accuracy and statistical performance that can *sometimes* arise from reliance only on exogenous variables.

We see a continued role for sewage treatment models that include an explanatory variable for treatment works size, especially given the statistical results for these models.

Nonetheless, stepping back and considering the full set of Ofwat consultation models, we consider that it would make sense for these to be *complemented* by some models that replace the treatment works size

[21] See for example Ofwat's comments on a variable suggested for the percentage of combined sewers in sewage collection models on page 45 of its consultation.

variable with the type of population density variable used by Ofwat for other parts of the value chain. Some of the key reasons are as follows:

- In terms of the underlying cost drivers affecting efficient levels of sewage treatment costs, what matters is the opportunities for economies of scale rather than the actual scale of water companies' sewage treatment works. Measures of population density are highly relevant to the differences between companies in these opportunities.
- Using an exogenous population density variable is more consistent with an outcomes-based regulatory approach than using an explanatory variable for treatment works size.
- A focus on the treatment works size variables poses risks of deterring a water company from making organisational changes, or planning capacity expansion, in a way that improves efficiency via increasing the load at treatment works.
- If models focus only on treatment works size variable, a company that takes decisions that enable greater economies of scale to be realised could be unfairly denied the benefits of those decisions relative to other companies.

Even if a model with an exogenous density variable does not perform as well statistically as a model that uses a treatment works size variable – e.g.in terms of the p-value on the explanatory variables or on metrics of modelled costs sensitivity and goodness of fit – we consider that such a model could play a useful role as part of a diverse suite of models with different benefits and drawbacks.

Q4.3: Do you agree with our approach to modelling population density? Which of the three explanatory variables do you support? (a) Weighted average density – LAD from MSOA; (b) Weighted average density – MSOA; (c) Properties per sewer length?

Neither agree nor disagree – N/A.

See our response to Q3.3 for some more general comments on the choice of density metrics. In that part of our response, we explain that we see the choice of density metrics to be an empirical matter in the light of model estimation results, rather than something that can be decided on an a priori basis. We have a general preference for exogenous density variables although there may be a role for endogenous metrics as part of the model suite if these are much better statistically.

For the six consultation models, we did not identify strong grounds to favour any one of the density metrics to the exclusion of the others. However, there may be a case for preferring the two exogenous density variables relative to the properties per sewer length variable.

Drawing on the enhanced metrics for modelled cost sensitivity and goodness of fit discussed earlier in this response, we found some mixed results:

- For models without the rainfall variable, the model with the "Weighted average density – MSOA" variable seemed to perform slightly better in terms of the metrics of goodness of fit and modelled cost sensitivity than the model with the "Weighted average density – LAD from MSOA" variable. But for the models with the rainfall variable, the model with the "Weighted average density – MSOA" variable seemed to perform slightly worse on modelled cost sensitivity and slightly better on goodness of fit.

- The models with properties per length of sewer perform worse in terms of modelled cost sensitivity than the models with MSOA-based metrics, but better in terms of goodness of fit.

Furthermore, using models with the exogenous explanatory variables for density might help make the modelling less vulnerable to cost interactions across the value chain. In contrast, using the total length of sewers in wastewater models, rather than an exogenous population density variable, may unfairly penalise companies that have less extensive sewer systems but greater costs in sewage treatment due to the operation of a larger number of smaller sites. By the same token, and as we discussed above, we see grounds for using an exogenous density variable for some of the sewage treatment models.

Q4.4: Do you agree with our proposal to assume a linear relationship between population density and sewage collection base costs?

Agree.

In the models that we submitted in January 2023, we included the squared weighted average density terms in our sewage collection model. However, we recognise that the scope of costs covered by our sewage collection model differed to that which Ofwat is now consulting on (e.g. we excluded growth and network reinforcement expenditure) and that some of Ofwat's models include the new rainfall variable which we had not included.

In the specific context of Ofwat's consultation models, we do not see a strong case at this stage for the inclusion of squared weighted average density terms in the sewage collection models.

We have found that, in some model specifications, the inclusion of squared weighted average density can enable the model to perform better than corresponding models that omit the squared variable. Whether there is statistical support for the squared term in specific sewage collection models may be affected by the new data for 2022/23 and by other aspects of the speciation of sewage collection models (e.g. choice of other cost drivers besides density). Given this, we do not have a definitive view at this point on whether a squared term should be used for PR24, but it seems reasonable to proceed without the squared term as a working assumption.

Q4.5: Do you agree with the inclusion of urban rainfall in our sewage collection and wastewater network plus models?

Agree.

On current evidence, we agree with the inclusion of the new urban rainfall variable in the set of sewage collection and wastewater network plus models.

We believe that the econometric models used for price control cost assessment should be subject to a process of continual improvement across successive price reviews. We welcome endeavours to increase the extent to which the inevitably simplistic econometric models capture the underlying cost drivers that lead to differences in the levels of efficient costs between companies.

For PR24 Ofwat has proposed a new urban rainfall variable defined as: "average rainfall falling in a company area (mm) multiplied by the urban company area (squared kms)". This draws on granular rainfall data provided by the Environment Agency and urbanisation data at the MSOA level of granularity.

The rationale that we see for the new urban rainfall variable is that it provides a measure of the scale of the surface water collection part of the wastewater service, which is particularly relevant to the sewage collection part of the value chain. While this variable is far from perfect it represents progress since PR19.

Wessex Water
Claverton Down
Bath BA2 7WW
Tel 01225 526 000
Web wessexwater.co.uk

Wessex Water Services Limited registered office
Registered in England No 2366648

We agree with Ofwat's interpretation that, for the aspects of statistical performance it has looked at, the Ofwat consultation models with the urban rainfall variable seem to produce good statistical results.

Drawing on Reckon's enhanced metrics of goodness of fit and modelled costs sensitivity, and looking at corresponding pairs of the consultation models: we found that:

- There were only very minor differences in performance between model SWC1 (without urban rainfall) and SWC4 (version with urban rainfall variable).
- Model SWC5 (with rainfall variable) performed better than SWC2 (without rainfall variable) in terms of both modelled cost sensitivity and goodness of fit.
- Model SWC56 (with rainfall variable) performed slightly better than SWC3 (without rainfall variable) in terms of goodness of fit and slightly worse in terms of modelled cost sensitivity.

These results suggest that while there is a good case for using the urban rainfall variable at PR24, its performance does depend on other aspects of model specification. Whether this new variable is used in all relevant models, or a subset of relevant models, for the final PR24 model suite is something that we suggest is decided in light of the evidence on model performance for the updated dataset and given other aspects of model specification.

Q4.6: Do you agree with our approach to capturing sewage treatment complexity in our proposed wastewater network plus base cost models?

Agree.

On the models and evidence currently available, we do not object to Ofwat's focus within its econometric model specifications on a single explanatory variable for treatment complexity concerning the proportion of load treated with ammonia consent ≤ 3mg/l. However, we consider that:

- The resulting models have serious shortcomings in relation to sewage treatment complexity and – more importantly – the environmental outcomes arising from companies' sewage treatment activities.
- Ofwat should take action outside of the econometric models to address the shortcomings of such models (e.g. off-model adjustments relating to the costs of P-removal).
- The models for sewage treatment should, on current evidence, include a time trend explanatory variable.

As with the corresponding PR19 models, Ofwat's consultation models do not take account of any differences between companies, and over time, in any aspect of these outcomes beyond that captured via a single explanatory variable for the proportion of load treated with ammonia consent ≤ 3mg/l. This approach is consistent with the sewage treatment models that we submitted in January 2023, which included the same ammonia consent as used by Ofwat.

Nonetheless, we consider it important to recognise that there are other important aspects of sewage treatment that are not captured by the 3mg/l ammonia consent variable. For instance, as reflected n the other explanatory variables that Ofwat said that it tried, these aspects may relate to discharge permits for:

- Phosphorous removal
- Biochemical Oxygen Demand.

- Ultra-Violet (UV) treatment.
- Ammonia, to the extent that this is not captured by the binary 3mg/l threshold.

As part of its overall cost assessment for PR24, Ofwat should keep these limitations firmly in mind. We were pleased to see that Ofwat is doing so in relation to P-removal as indicated by its follow-on question about the remuneration of the additional ongoing cost associated with P-removal (which we turn to in the sub-section that follows).

Furthermore, in the absence of econometric models that properly capture sewage treatment complexity and the associated environmental outcomes, we consider that it is particularly important for Ofwat to properly assess whether the evidence supports a time trend explanatory variable.

If there have been industry-wide increases in efficient levels of sewage treatment costs, driven by enhanced treatment complexity requirements and environmental improvements, which are not fully captured by the 3mg/l ammonia consent variable, then these would tend to be reflected in the estimation results for models that include a time trend variable.

We found that the evidence does indeed support a time trend explanatory variable for sewage treatment. We discuss this in more detail in Appendix A1.

Q4.6 (continued): What are you views on our proposed options to account for additional ongoing cost associated with P-removal? (a.) Models with a P-driver (eg percentage of load with a P-permit <= 0.5mg/l) fixed at the 2024/25 level; (b) a post-modelling adjustment that funds efficient ongoing opex associated with P-removal using data provided by companies in APRs; (c) Cost adjustment claims.

Of the three options presented by Ofwat, we support option (b): a post-modelling adjustment that funds efficient ongoing opex associated with P-removal using data provided by companies in APRs.

If it was feasible to properly capture the ongoing impacts of P-removal requirements on companies' efficient levels of costs via the econometric modelling (i.e. option (a)), we would favour this approach. But this does not seem likely to be feasible for PR24, based on Ofwat's findings and our own modelling analysis.

Option (b) offers a practical approach at PR24. Where important cost drivers are not captured in the econometric models, we see value in post modelling adjustments as a practical means to mitigate the deficiencies of the models. There is precedent from PR19 for the use of post modelling adjustments to help address shortcomings in the scope of underlying cost drivers captured in Ofwat's econometric models (the post-model adjustment for forecast population growth).

We would be keen to engage further with Ofwat as it develops its approach in this area. At this point we highlight the following:

- The ongoing cost associated with P-removal would tend to include both operating expenditure and capital maintenance expenditure.
- The efficient level of ongoing costs (relative to the scale of P-removal) could vary between companies depending on the types of enhancement solutions that they introduced (and were funded for) in the past. For instance, companies that adopted enhancement solutions with a greater emphasis on opex may, for quite legitimate reasons, have higher levels of ongoing costs than those which used and were funded for capex-intensive enhancements.

We do not consider that the cost adjustment process under Ofwat's option (c) is suitable at all. Given its evidential and process requirements, the materiality thresholds, and Ofwat's track record of allowing few cost adjustment claims relating to base costs, we do not how the cost adjustment process can provide a credible and effective way to address the underlying concern identified by Ofwat that the additional ongoing cost associated with more stringent phosphorus removal programmes across the sector may not be fully captured in Ofwat's proposed base cost models.

Furthermore, Option (c) risks entrenching, rather than helping to reduce, the capex bias arising from Ofwat's broader approach to enhancements.

Q4.7: Do you agree with Southern Water's proposal to include the percentage of population living in coastal areas in sewage treatment models?

Neither agree nor disagree – N/A.

On the evidence of Ofwat's consultation models, we see some support for including some sewage treatment models with the explanatory variable for the percentage of population living in coastal areas in the model suite, alongside some that do not include this variable.

We do not consider the evidence available at this stage to be sufficient to support exclusive reliance on this variable, but we think that it could play a role as part of a diverse suite of models with different strengths and weaknesses.

Our analysis of Ofwat's models, using Reckon's enhance metrics of goodness of fit and modelled cost sensitivity, is that the models with the coastal population variable have better goodness of fit and worse modelled cost sensitivity. The difference in performance depends on other aspects of model specification. For instance, for model SWT2, the modelled cost sensitivity is about 5% worse in the version with the coastal population variable whereas for model SWT1 the modelled cost sensitivity is about 35% worse in the version with the coastal population variable.

In its consultation, Ofwat said that the estimated coefficient on the coastal variable is sensitive to dropping Southern Water from the dataset, and that its analysis suggested that this variable may be capturing a Southern Water specific impact, rather than an overall industry-wide impact of operating in coastal areas.

We feel that this finding may reveal more about the limitations of Ofwat's current approach to looking at sensitivity than it does about the suitability of the coastal population variable.

We carried out a more systematic assessment of sensitivity, looking at how the estimated coefficient for the coastal population variable varied when dropping specific companies and years from the dataset. We found that for model SWT1 the coefficient did turn negative for some dataset variations. But for model SWT2 the coefficients were positive for all datasets that we tried. We summarise the results from this exercise in the charts below. These results are consistent with the finding above that modelled cost sensitivity is considerably worse for model SWT1 when the coastal population variable is applied, but the impact is smaller for SWT2 and SWT3.

Wessex Water
Claverton Down
Bath BA2 7WW
Tel 01225 526 000
Web wessexwater.co.uk

Wessex Water Services Limited registered office
Registered in England No 2366648

Figure 6 Sensitivity of coefficient on coastal population variable to minor dataset variations: SWT1

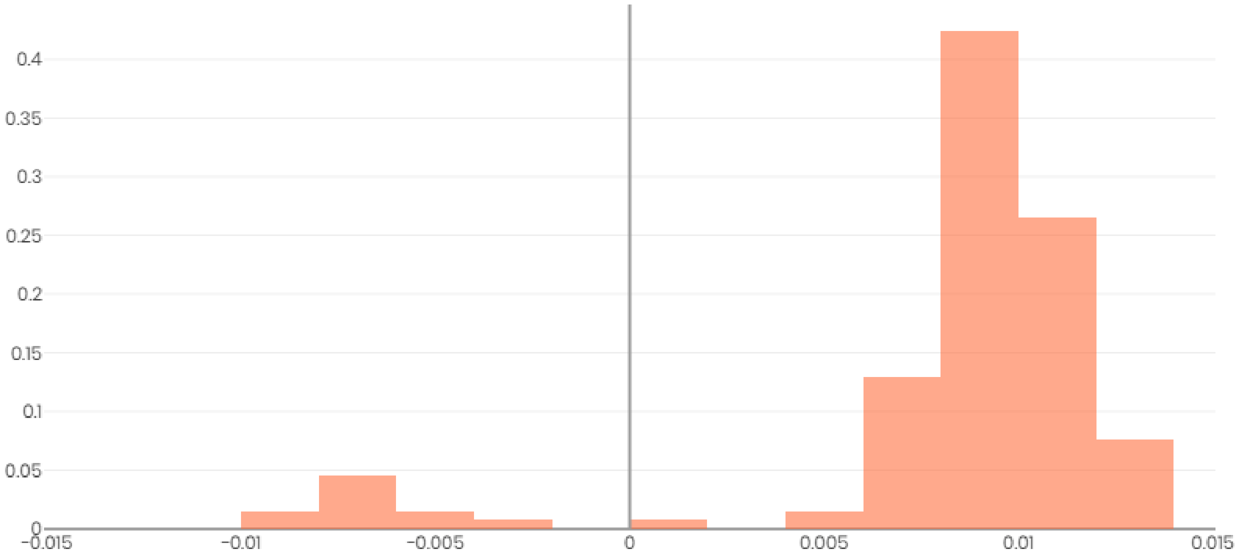
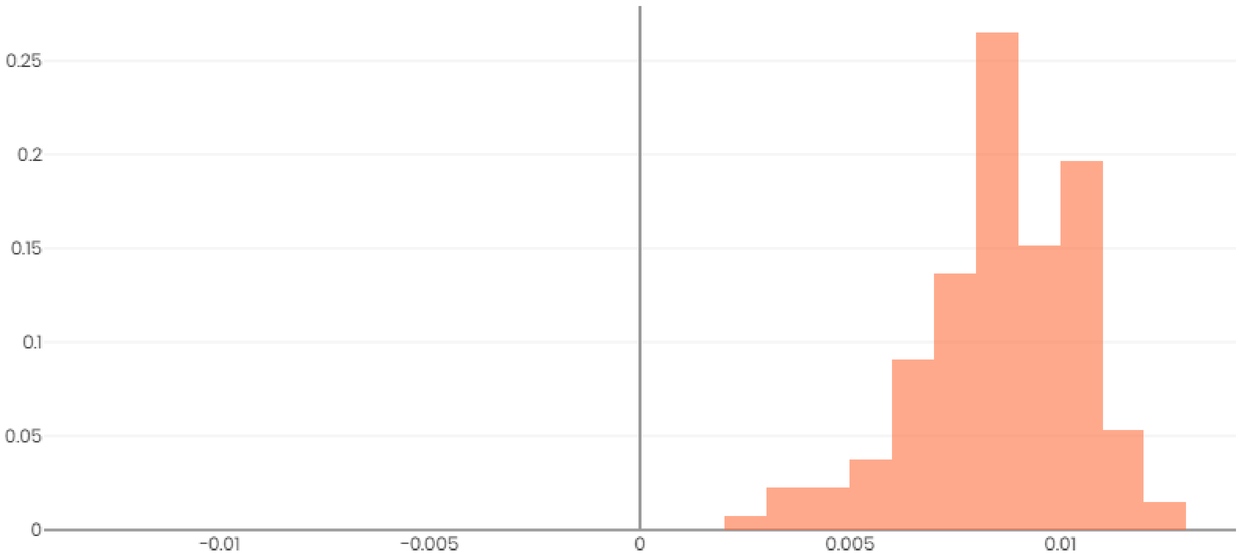


Figure 7 Sensitivity of coefficient on coastal population variable to minor dataset variations: SWT2



Q5.1: Do you agree with our proposed set of bioresources cost models?

Strongly disagree.

There are a number of elements of the bioresources cost models that we support. However, we do not consider the set of consultation models to be suitable for PR24 as they stand.

Ofwat supports its models by reference to the procedures used to develop and test them. However, as discussed in detail in Appendix A1, we see significant limitations in Ofwat's approach to model robustness assessment. This seems to (a) give too much emphasis to pass/fail tests that are not particularly useful for the purposes of assessing the relative performance of alternative models under consideration and (b) miss out on good ways to capture the most important aspects of model robustness and model performance. Improvements to this aspect of Ofwat's approach will be important as Ofwat reflects on stakeholder responses to this consultation and if it looks to narrow down its set of models.

As discussed in Appendix A2, we disagree with Ofwat's rejection of time trends as a matter of principle.

The points we make about density variables and exogenous variables also apply to bioresources (e.g. see discussion on MSOA-related metrics in our response on wholesale water and on economies of scale for sewage treatment).

The range of efficiency scores implied by Ofwat's bioresources models is quite wide. This raises questions about whether a full enough set of cost drivers has been covered.

For one of our two bioresources models submitted in January 2023, we included an explanatory variable relating to the load received at STWs subject to Secondary Activated sludge treatment. This had performed well in our modelling. Ofwat does not to have properly responded to this suggestion its modelling (commenting instead on the results from bioresources models with phosphorus and ammonia consents).

We consider that it will be important to use a diverse set of models at PR24, reflecting a range of alternative explanatory variables.

We are concerned about the inclusion of growth-related expenditure within the base models, without further information from Ofwat as to how it will adjust for differences between companies and over time in the drivers of such expenditure.

See also our comments on enhancement operating expenditure under question 3.1.

Q5.2: Do you agree we should use unit cost models to assess bioresources expenditure?

Agree.

We would agree with Ofwat using the unit cost models for bioresources at PR24, and not using the aggregate cost models at all for bioresources.

Indeed, we see quite clear evidence to favour the unit cost models, at least within the set of model variants that Ofwat has consulted on.

In terms of the conventional model estimation results reported by Ofwat, we highlight the following:

- The p values on the density explanatory variables for density, economies of scale in sludge treatment, and location of sewage treatment works relative to sludge treatment centres, seem to be lower (i.e. better) in the unit cost models compared to the aggregate cost models when comparisons are made between models that have the same density variables (e.g. BR1 unit cost versus BR4 aggregate cost).
- There seem to be counter-intuitive results in Ofwat's aggregate cost models for bioresources. While Ofwat reports that "all models are consistent with engineering, operational and economic rationale" and that "all estimated coefficients on the explanatory variables are of the expected sign and plausible

magnitude" we would question the results from the aggregate cost models. The coefficients on the sludge produced scale variable are all over one. This implies a form of *company-wide diseconomies of scale*. We do not think that there is an engineering, operational and economic rationale for this finding. We do not see any reason why a company that has twice the scale in terms of sludge produced as another company, but the same position in terms of density and other explanatory variables (e.g. density) would have more than twice the costs if acting efficiently. The unit cost models do not suffer from this problem.

- As Ofwat rightly observes, the lower R-squared for the unit cost model is simply a result of the way that these models are specified. It is not a reliable guide of goodness of fit compared to aggregate cost models.

Furthermore, we carried out an analysis using Reckon's enhanced metrics of goodness of fit and modelled cost sensitivity (introduced in Appendix A1). To assess this on a like-for-like basis, we compared the four unit cost models from Ofwat's consultation (BR1u, BR2u, BR3u, and BR4u)[22] against the corresponding aggregate cost models which have the same explanatory variables besides the scale variable (i.e. BR4a, BR5a, BR6a and BR3a respectively),

We found evidence that the unit cost models performed consistently better in terms of modelled cost sensitivity, without any material loss of goodness of fit:

- Model BR1u had better performance in modelled cost sensitivity and very similar goodness of fit relative to model BR4a.
- Model BR2u had better performance in modelled cost sensitivity and very similar goodness of fit relative to model BR5a.
- Model BR3u had better performance in modelled cost sensitivity and very similar goodness of fit relative to model BR6a.
- Model BR4u had better performance in modelled cost sensitivity and very similar goodness of fit relative to model BR3a.

Q6.1: Do you agree with our proposed set of residential retail cost models?

Strongly disagree.

We disagree with Ofwat's residential retail models as they stand.

One fundamental problem with Ofwat's models is the omission of time trend explanatory variables, contrary to the evidence (see discussion in Appendix A2). This in turn has had wider adverse effects on Ofwat's models:

- Losing the explanatory variable for the meter penetration cost driver, which represents a step backwards compared to the models used at PR19.
- Picking the wrong explanatory variables for what Ofwat calls the "Propensity to default on water bill payment".

[22] We added the "u" suffix here to differentiate from the aggregate cost models which we denote with an "a" suffix.

Addressing these points would make a substantial improvement to the residential retail benchmarking models.

We welcome Ofwat's use of time dummy variables for Covid-19, as we had suggested in our January submission. However, the evidence suggests that these dummies are not needed for the models of other retail costs, once a time trend is included. Furthermore, these dummies are needed alongside a time trend for bad debt related costs. There is a possible case for a dummy for 2021/22; we suggest that this is reconsidered in light of models run when new data up to 2022/23 becomes available.

In our submission from January 2023, some of our models of bad debt related costs were models in which the dependent variable was the ratio of bad debt related costs to total billed revenue. In its consultation, Ofwat said that it was not clear why this approach would be better than the unit cost modelling approach applied in PR19. We were a little surprised by this response. There is an intuitive logic and precedent for benchmarking companies on the basis of bad debt as a proportion of revenue. Indeed, this is what PwC did extensively in its report for Ofwat on residential retail costs in 2022.[23] Furthermore, the validity of this approach for the econometric models of bad debt related costs is an empirical question. There seems to be some evidence in favour of this approach. For example, in its report for us Reckon found that there tended to be a narrower range of efficiency scores under this approach.[24]

We agree with the use of an explanatory variable for the average bill size. However, it is important that this reflects the average bills faced by each company's customers and the revenue at risk. In its work for us on benchmarking residential retail costs, Reckon identified an issue in the case of the average bill size variable for South West Water. This concerns the £50 per customer contribution from Government that South West Water's customers benefit from. For its modelling, Reckon made an adjustment to the average bill size variable to reflect the estimated impact of the customer contribution on average bill size and revenue at risk. We suggest that Ofwat considers this issue further in light of the adjustment used by Reckon.[25]

We agree with modelling both at the disaggregated level (the RDC and ROC models) and aggregated models (the RTC models). There is some evidence of the total cost models being inferior (e.g. loss of explanatory variables that are used at the disaggregated level or lower p-values). Further consideration will be needed, ahead of draft determinations, to decide on what weight, if any, to give to the *aggregated models*. The enhanced metrics of modelled cost sensitivity and normalised goodness of fit discussed in Appendix A1 can help inform on this question because they can be calculated for groups of triangulated models rather than just for individual modes.

A further issue on the modelling of residential retail costs concerns the **smoothing of depreciation**. Ofwat states that rather than using depreciation data reported by companies it has replaced this with a measure of smoothed depreciation which is calculated by taking the company average of the historical depreciation data provided by companies over the full length of the dataset (2013/14 – 2021/22).[26] This is a bizarre and completely indefensible approach. Ofwat now finds itself in the position of benchmarking wholesale activities using comparisons of cash expenditure between companies (with no smoothing over time of capital expenditure) yet for residential retail activities it is smoothing over a 9-year period the reported data

---

[23] PwC (2022) *Retail services efficiency review 2022*.

[24] Reckon (2023) *Residential retail cost assessment at PR24: econometric benchmarking models*, page 13, https://www.reckon.co.uk/s/2023-04-06-Reckon-Benchmarking-models.pdf.

[25] Reckon (2023) *Residential retail cost assessment at PR24: econometric benchmarking models*, page 78, https://www.reckon.co.uk/s/2023-04-06-Reckon-Benchmarking-models.pdf.

[26] Ofwat (2023) *Econometric base cost models for PR24*, page 59.

on depreciation (despite the depreciation figures being already smoothed, based on asset life assumptions that companies use for accounting purposes).

Q6.2: Do you agree with our approach to modelling deprivation, and/or have any views on the selected variables?

Strongly disagree.

We agree with the inclusion of models that use an income deprivation score variable. We also agree with the approach of looking for other variables, such as those from Equifax, to complement the income deprivation score variable, to avoid reliance on any single variable.

However, the evidence suggests that Ofwat's model development process has been undermined in this area by not considering time trend variables.

We show in the table below estimation results for the bad debt related models from Ofwat's consultation, together with variants on these that include a time trend and/or the alternative Equifax variable from Ofwat's dataset which did not feature in its consultation models (eq_rgc102).

The key points we draw from the table are as follows:

- When a time trend is added to the Ofwat models (RD1, RD2, RDC3), the time trend is significant at the 5% or 10% level across all three Ofwat model variants.
- For the Ofwat model with the income deprivation score, this variable is significant at the 5% level even with the inclusion of the time trend (model RDC3tt).
- However, for Ofwat's other two models, the Equifax variables are not significant at 10% when a time trend is included (models RDC1tt and RDC3tt). The case for favouring these variables is highly questionable.
- The alternative Equifax variable eq_rgc102 is not significant in the absence of a time trend (model ALT1). But, once a time trend is included, the eq_rgc102 variable is significant at the 10% level and almost at the 5% level (ALT2).
- In addition to the models included in the table we also tested the effect of including a time dummy for 2021/22 alongside the 2018/19 and 2019/20 dummies included in the table models. Under these conditions the variable eq_rgc102 becomes significant at the 5% level.

Wessex Water
Claverton Down
Bath BA2 7WW
Tel 01225 526 000
Web wessexwater.co.uk

Table 6 Estimation results for Ofwat models of bad debt related costs and some variants of these

Model Ref	RDC1	RDC2	RDC3	RDC1tt	RDC2tt	RDC3tt	ALT1	ALT1tt
Model type	Ofwat consultation model			Ofwat model + time trend			Models with alternative Equifax variable	
Dependent variable	Ln(Bad debt related costs per household, CPIH adjusted)							
Explanatory variables								
Ln (revenue per household)	1.170*** (0.000)	1.207*** (0.000)	1.045*** (0.000)	1.125*** (0.000)	1.137*** (0.000)	1.035*** (0.000)	1.248*** (0.000)	1.085*** (0.000)
eq_lpcf62	0.064*** (0.007)			0.034 (0.101)				
eq_xpcf2		0.879** (0.017)			0.407 (0.373)			
Credit risk score (eq_rgc102)							-0.014 (0.352)	-0.029* (0.059)
Income deprivation score			0.089*** (0.002)			0.057** (0.023)		
Time trend				-0.033** (0.035)	-0.037* (0.065)	-0.034** (0.026)		-0.050*** (0.004)
Time dummy variables?	2019/20 & 2020/21	2019/20 & 2020/21	2019/20 & 2020/21	2019/20 & 2020/21	2019/20 & 2020/21	2019/20 & 2020/21	2019/20 & 2020/21	2019/20 & 2020/21
Overall R-squared	0.66	0.66	0.68	0.67	0.66	0.69	0.64	0.68
Observations	153	153	153	153	153	153	153	153

Given the analysis and evidence above, our proposed approach is as follows:

- All models of bad debt related costs to have time trends as well as Covid-19 dummy variables.
- Some models of bad debt related costs to use the income deprivation score variable and some to use the Equifax variable eq_rgc102.
- No use of the variables eq_lpcf62 and eq_xpcf2.
- Corresponding updates to the models of total retail costs.

Wessex Water
Claverton Down
Bath BA2 7WW
Tel 01225 526 000
Web wessexwater.co.uk

Wessex Water Services Limited registered office
Registered in England No 2366648

Q6.3: Do you agree with the inclusion of Covid-19 dummy variables in the residential retail cost models?

Agree.

For models of bad debt related costs, we see a strong case for the inclusion of Covid-19 dummy variables. This is regardless of whether a time trend variable is, as we propose elsewhere, included in these models.

There is a question of whether a Covid-19 dummy variable is appropriate for 2021/2022 as well as for 2019/20 and 2020/21. It is far from clear that the distortions to industry-wide bad debt costs associated with the Covid-19 pandemic are limited to 2019/20 and 2020/21. In the models we submitted to Ofwat in January 2023, we used three dummy variables. There will be a better evidence base to decide whether to apply two or three dummy variables for the effects of the Covid-19 pandemic once data is available for 2022/2023 for inclusion in the econometric models.

For models of other retail costs (i.e. costs that exclude bad-debt related costs), the underlying dynamics of industry-wide costs over time seem to be much better captured by a time trend. Once that correction is made, we would be unsure if there is a need for Covid-19 dummy variables - both on statistical grounds and in terms of the operational and economic rationale.

Q6.4: Do you agree with the removal of transience from the residential retail cost models?

Agree.

We agree with the removal of transience from the residential retail cost models.

The analysis we commissioned from Reckon did not support the use of the variable used at PR19 to proxy for transience. Some points to highlight from Reckon's work are as follows:

- While the variable used at PR19 was intended to capture customer transience or migration, it was not a direct measure of this. It was based on ONS data relating to migration flows in and out of local authority districts (LADs) which were used to construct a metric that proxied customer transiency. The ONS data on migration flows do not reflect household moves within an LAD, which could account for a significant share of population flows. This means that the PR19 variable could provide a distorted picture of differences in customer transience across companies.
- Reckon found that, across a range of different model specifications, the estimated coefficient on the variable tended to have low t-statistics and in some of the model variants, the sign was negative, which is counter-intuitive. This seems consistent with what Ofwat found.

Q6.5: Do you agree with the removal of 'proportion of metered customers' from the residential retail cost models?

Strongly disagree.

We disagree with the removal of 'proportion of metered customers' from the residential retail cost models.

As evidenced in Appendix A2, the loss of the explanatory variable for the proportion of metered customers seems to be an inadvertent side effect of Ofwat's decision not to include time trend variables in its models of residential retail costs, despite the strong evidence to support the use of such variables.

Once a time trend variable is allowed for, the proportion of metered customers explanatory variable is supported by the types of evidence that Ofwat gives weight to (i.e. economic rationale and the p-value on the explanatory variable for the proportion of metered customers).

Wessex Water
Claverton Down
Bath BA2 7WW
Tel 01225 526 000
Web wessexwater.co.uk