love every drop
anglianwater

PR24
Base cost modelling and response to companies' symmetrical cost adjustment claims

Anglian Water
October 2022

Appendix – Comments on base cost modelling and response to companies' symmetrical cost adjustment claims

Introduction

This appendix to the Business Plan covers the base cost models and symmetric cost adjustment claims (CACs). Where appropriate, we provide -

- further comments on the proposed base cost models (in addition to what we already said in our May 2023 base cost consultation response)
- further support to the CACs we submitted in June 2023, and
- responses to the symmetrical cost adjustment claims made in June by other companies.

In order to benefit from all the information available to date, our conclusions incorporate the impact of the inclusion of the 2022/23 data in the modelling, which will be the latest year considered at the draft determinations.

First, we cover the water modelling by considering water network topography, United Utilities' CAC on reservoirs maintenance and metering costs (with an overlap with the retail modelling for the latter).

Second, we provide our thoughts on the wastewater modelling by focusing on economies of scale at water recycling centres, growth costs at water recycling centres as well as urban rainfall alongside the two CACs submitted by Yorkshire and United Utilities on combined sewers and drainage.

Finally, we study the three density variables proposed by Ofwat and reiterate our view on the most suitable density driver.

I. Network topography

This section contains our comments on companies' responses to the base cost consultation on the inclusion of average pumping head (APH) in the water models as well as the three cost adjustment claims (CACs) that have been submitted on this topic, namely by South Staffs (SSC), Severn Trent (SVE) and SES Water (SES). Based on both the engineering rationale and the statistical evidence, this section also reiterates our position on the best way to capture network topography in the modelling.

Contents

1. Summary of our position

Reprising the arguments developed in our May base cost consultation response, our APH CAC and in this document, we summarise below our position on the topography question. This is ordered by degree of importance for PR24, with the first two points being critical to ensure an accurate assessment of the network topography and its associated underlying costs.

1. The APH attributable to treated water distribution (TWD) should be included in TWD and wholesale models to capture the differences between companies in network topography and associated energy requirements.
2. Booster pumping stations (BPS) per length of mains should be removed from the modelling suite since it is an inappropriate driver to capture network topography.
3. There appears to be merit in considering the APH attributable to water resources plus (WRP) and total APH in at least a subset of models to complement the inclusion of APH TWD.
4. If Ofwat were to keep relying in some way on BPS in the modelling (which we contest), it would be much more appropriate to consider the associated capacity rather than simply counting the number of BPS[1].
5. If Ofwat were to keep relying in some way on BPS in the modelling (which, as we have said, we contest), it would be inappropriate to include both APH and BPS (whether the total number or the associated capacity) within the same model since they are supposed to capture the same effect.

[1] This would align with the position in wastewater modelling where the associated capacity/actual volume of load treated is considered in the modelling, not simply the number of pumping stations.

We provide additional evidence to explain our position in the subsequent sections.

2. Base cost consultation

We note that the majority of the industry supports the inclusion of APH in the modelling suite as only six companies have expressed concerns about its use, namely Affinity (AFW), Northumbrian (NES), Portsmouth (PRT), Southern (SRN), United Utilities (UU) and Welsh (WSH).

The aim of this section is not to reformulate all the different arguments that we have made in our May base cost consultation response or in our cost adjustment claim, but rather to focus on the main arguments developed by the industry on the best way to capture network topography in the models.

While stakeholders generally acknowledged the strong engineering rationale of APH to capture network topography and energy requirements, the main arguments from those companies opposing its inclusion in the PR24 modelling suite related to data quality and the presumed low reliability of the historical data which would alter the robustness of econometric models. We summarise below the main arguments that have been made in this respect and study them all together in our response.[2]

a. Data quality of APH TWD is poor for at least 11 out of the 12 years (AFW and UU)[3]
b. If intended to reflect topography, APH TWD should not vary significantly across years (AFW and NES)
c. The better significance of the APH models may be spurious (UU and PRT)
d. Econometric models are not more robust with APH (AFW).

To address the above concerns on the alleged poor data quality of APH TWD over 2011/12 - 2020/21 that would alter the statistical performance of the econometric models, **we have run two different sensitivities to show that this is not the case and that the estimated relationship between APH and TWD and wholesale costs is not spurious.**

AFW presented analysis which showed a handful of outliers where the range of their reported value for APH TWD has varied substantially between 2011/12 and 2020/21. Starting from this, we have carried out two types of analysis where in each case we have updated the modelling with the 2022/23 data. The data from this year are considered to be of better quality as they fully account for Ofwat's updated guidance.[4][5]

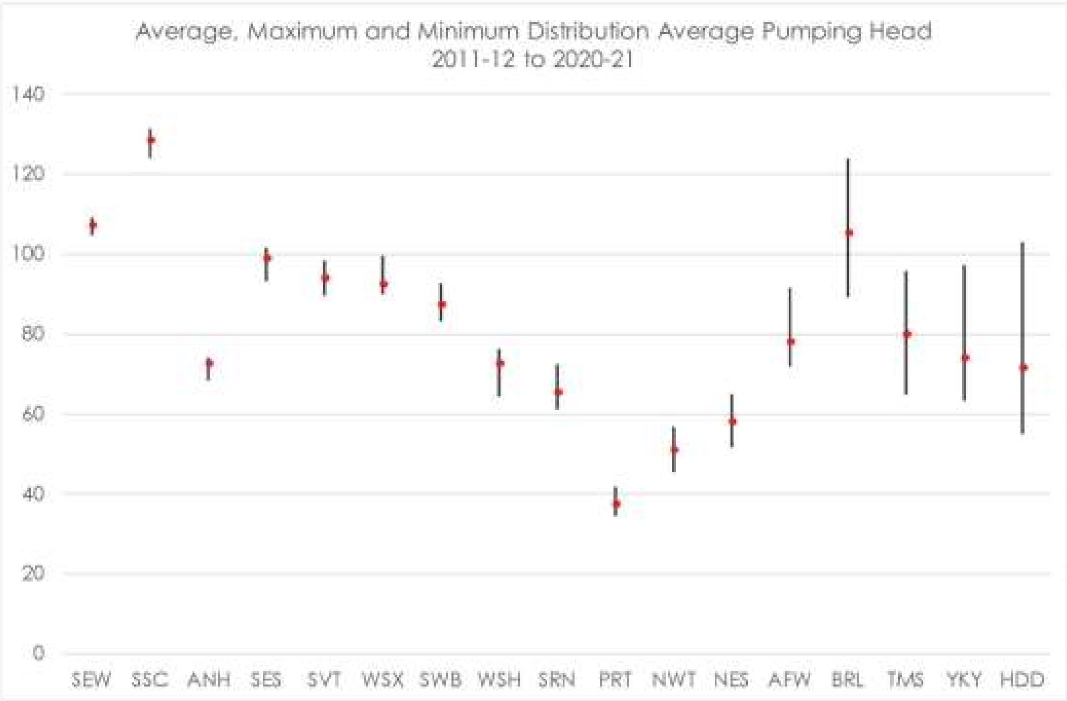
- Sensitivity 1—Re-run the PR24 modelling suite by replacing the APH TWD values over 2011/12 - 2021/22 with the 2022/23 value.

[2] At least one company is quoted for each of these arguments.
[3] Since this response was developed before the publication of the 2022/23 data, this argument is now equivalent to say that the data quality if poor for at least 11 out of the 13 years.
[4] The conclusions remain the same if we do not replace the 2021/22 value (which already accounts for Ofwat's updated guidance). We have kept unchanged the values of SVT, SWT, BWH and DVW as the 2022/23 data cannot be robustly back-casted for the pre-merger period.
[5] https://www.ofwat.gov.uk/wp-content/uploads/2022/05/Average-Pumping-Head-Data-Quality-Improvement-Final-Report-.pdf

- Sensitivity 2—Same as above but only replacing the APH TWD values for the five outliers identified by AFW (i.e. HDD, YKY, TMS, BRL and AFW) as illustrated in the chart below.

Given that APH is a proxy measure for topography, it is reasonable to assume the value remains broadly constant over the period modelled.

Figure 2.1 Maximum range of APH TWD at a company level, 2011/12-2020/21



Note: A similar graph has been shown by NES on p. 5 of its consultation response but with an exclusive focus on WaSCs. The same results apply since the underlying data used by AFW and NES are the same.

Source: Affinity Water (2023), 'Econometric Base Cost Models for PR24 – Consultation Response', p. 2.

In both cases we find that the estimated coefficient of APH TWD is highly statistically significant in 17 out of the 18 models tested which confirms the key importance of APH TWD in the modelling and the sufficient reliability of the historical data.[6] Replacing the historical APH TWD data with the 2023 values does not change the underlying efficiency results, since the vast majority of the estimated company efficiency rankings remain unchanged. The correlation in terms of ranking between modelling with unadjusted APH data and modelling with adjusted APH data, as described above, is 0.99.[7]

Table 2.1 Statistical performance of APH TWD in TWD models

[6] The p-value is 0.13 in the only case where APH TWD is not statistically significant at a 10%, 5% or 1% level which, in line with Ofwat's PR24 approach, still translates to strong enough statistical performance for its inclusion in the modelling.
[7] In both cases, whether we replace the APH TWD values for the whole industry or just for outliers.

	Model 1	Model 2	Model 3
Sensitivity 1	0.380***	0.472***	0.404***
Sensitivity 2	0.391***	0.469***	0.398***
Unadjusted APH data	0.343***	0.400***	0.348***

Note: *** Statistically significant at the 1% level, ** at the 5% level, * at the 10% level. A log-transformation of APH TWD is used in each case.

Source: Anglian analysis from Ofwat's PR24 dataset and the 2022/23 APR.

Table 2.2 Statistical performance of APH TWD in WW models

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Sensitivity 1	0.351***	0.322**	0.409***	0.373***	0.277*	0.245(0.131)
Sensitivity 2	0.383***	0.361***	0.423***	0.402***	0.306**	0.280*
Unadjusted APH data	0.330***	0.317***	0.342***	0.330***	0.272**	0.256**

Note: *** Statistically significant at the 1% level, ** at the 5% level, * at the 10% level. A log-transformation of APH TWD is used in each case. When the estimated coefficient is not statistically at a level of at least 10%, the p-value is provided in brackets.

Source: Anglian analysis from Ofwat's PR24 dataset and the 2022/23 APR.

Regarding the relative performance of APH TWD and BPS in the models, we note that, although it was broadly similar during the consultation phase with data up to 2021/22, the statistical performance of BPS has deteriorated relative to that of APH TWD with the inclusion of the 2022/23 data.

Indeed, while APH TWD is always statistically significant at the 1 percent level in wholesale models with both WAD measures, this is the case for BPS only in the models with WAD MSOA; it is statistically significant at the 5 percent level only in the models with WAD LAD from MSOA. The trend is the same with properties/mains: APH TWD is more statistically significant than BPS in these models (at the 5 percent level for APH TWD, and only at the 10 percent level for BPS in one model).

Based on our analysis, we conclude the following:

- Although we acknowledge that the historical APH TWD data before 2021/22 may not be as robust as for 2021/22 or 2022/23 since they are not fully based on Ofwat's updated reporting guidance, **we show that the historical period can still be robustly relied upon**. Indeed, when giving a 100 percent weight to the latest 2022/23 APH TWD values **we demonstrate that the econometric models still perform as strongly**. However, if Ofwat were to judge it as a necessary adjustment, it would be legitimate to replace a few atypical (and potentially unjustified) APH TWD values without undermining the robustness of the models.

- **The estimated relationship between APH TWD and both the TWD and wholesale costs is not spurious but rather reflects the strong engineering rationale of APH as a key topography driver**, which is recognised by the whole industry.

- One could even say that if more weight were to be given to the 2022/23 value, the relationship between APH and costs would even be stronger since the estimated coefficient is almost always higher than in the baseline scenario with unadjusted APH data.

- The wholesale econometric models perform better with APH TWD than with BPS (see Section 3 for the associated statistical results).

To conclude this section, which focuses on APH TWD versus BPS, with the base cost modelling response as a starting point to analyse the different arguments to date, it is worth (re)summarising the numerous weaknesses and inconsistencies of BPS. The points below clearly indicate that BPS is *not* a topography driver and therefore not suited for PR24 since we have available a much stronger alternative with APH TWD.

i. BPS is completely uncorrelated with energy requirements related to TWD (R-squared of 0.02 of model of TWD power costs against BPS as indicated in page 8 of the appendix of our base cost modelling response and consistent with the low R-squared we found during our CMA appeal).[8]
ii. While the whole industry acknowledges that APH TWD is a topography driver, SSC shows in its CAC that BPS per length of mains is completely uncorrelated to APH TWD (R-squared of 0.066, page 13). This clearly demonstrates again that BPS is not a topography driver.
iii. Unlike APH TWD, BPS is partly correlated with density variables (c. 0.6), raising concerns about overfitting companies that operate in dense areas (page 14 of SSC's CAC).
iv. Econometric wholesale models with APH TWD consistently perform better than wholesale models with BPS.
v. The economic intuition of counting the number of BPS regardless of their associated capacity is non-existent. This is perfectly illustrated by SSC in page 6 of its CAC with two BPS of a completely different capacity, 12MW vs 30kW, being currently considered as having exactly the same energy or maintenance requirements. We develop this point further in Section 4 dedicated to SSC's CAC.

Finally, we would also like to note that we agree with a point made by SES in its base cost consultation response : '*If mains length increases below the necessity for building an extra booster pumping station (meaning the number of boosters remains the same while the mains length keeps growing), the value of booster per km of mains will decline – but this should not be taken to mean that the overall cost of operating the network has in fact fallen*' (Q3.2).

This is indeed true and represents another issue faced by BPS. We note that, by definition, this issue is not present when relying on APH TWD instead.

While the focus on this section has been on the base cost consultation and therefore on Ofwat's current modelling suite, we cover additional points below in the relevant CACs sections such as the use of APH WRP and total APH in addition to APH TWD in TWD models or the joint use of BPS and APH in the same models.

3. SES' CAC on pumping costs for wholesale water

[8] Anglian Water (2020), 'PR19 CMA Redetermination Response to Provisional Findings', p. 16.

While the methodology proposed by SES departs from Ofwat's proposed modelling suite in terms of the dependent and independent variables used, we fully agree on the necessity to remove BPS from the modelling suite for the reasons stated in Section 2 above. In addition to this key point, SES proposes to include the remaining component of total APH in the models, i.e. APH WRP. We agree on the latter as well, at least for a subset of models. In our May base cost consultation response, we focused on APH TWD versus BPS but since then **the superiority of APH TWD over BPS has been clearly demonstrated, our view is that Ofwat and the industry should now focus on the relevance of APH WRP and total APH in the models as a complement to APH TWD**. This could further reinforce the robustness of the current wholesale water (WW) models with APH TWD.

Indeed, the engineering rationale for APH WRP is strong and recognised by Ofwat and the industry. This is particularly relevant for borehole abstraction where the relevance of pumping head is obvious. We understand that the reason why Ofwat did not propose APH WRP in the consultation phase was due to its limited performance in the models : '*We tested WRP APH (water resources plus raw water distribution plus water treatment APH) in WRP models, but found it is not a significant driver of WRP costs*' (page 24 of the consultation document).

However, we find that the inclusion of the 2022/23 data partly alleviates the issue as the statistical performance of APH WRP improves in WRP models and the statistical performance of total APH is very high and even generally better than APH TWD in the WW models.

The statistical results of APH WRP are presented in Figure 3.1 below. To put this statistical performance in perspective with the weighted average composite (WAC) treatment variable included in Ofwat's WRP proposed models, the average p-value of APH WRP in this case, 0.15, is better than the average p-value of WAC at the time of the consultation, 0.19. Consequently, we commend Ofwat to consider the inclusion of APH WRP in a subset of WRP models.

Table 3.1 Statistical performance of APH WRP in WRP models

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Ln(APH WRP)	0.113(0.225)	0.111(0.160)	0.126(0.179)	0.120(0.134)	0.57(0.115)	0.151* (0.078)

Note: * Statistically significant at the 10% level. P-value of the estimated coefficient is provided in paratheses.

Source: Anglian analysis from Ofwat's PR24 dataset and the 2022/23 APR.

We note that, unlike APH TWD, total APH is always statistically significant at the 1 percent level in the wholesale models (see Table 3.2). While the statistical performance of total APH is marginally lower than APH TWD on the models relying on WAD measures (models 1-4), the consideration of total APH (rather than APH TWD) alongside properties/mains (models 5-6) represents an improvement. Since we have shown in Section 2 that APH TWD always performs better than BPS, the evidence below increases the difference in statistical significance between BPS and APH.

Table 3.2 Statistical performance of APH TWD and total APH in wholesale models

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Ln(Total APH)	0.312***(0.006)	0.290***(0.006)	0.336***(0.004)	0.314***(0.004)	0.361***(0.001)	0.339***(0.001)
Ln(APH TWD)	0.330***(0.002)	0.317***(0.002)	0.342***(0.003)	0.330***(0.004)	0.272**(0.024)	0.256**(0.035)
Ln(BPS/mains)	0.406**(0.019)	0.407**(0.011)	0.466***(0.006)	0.457***(0.005)	0.341*(0.052)	0.326**(0.049)

Note: *** Statistically significant at the 1% level, ** at the 5% level. [1] A positive number means than total APH TWD is more statistically significant than total APH, and vice-versa. P-value of the estimated coefficient is provided in brackets.

Source: Anglian analysis from Ofwat's PR24 dataset and the 2022/23 APR.

To conclude, **there is sufficient evidence to use total APH instead of APH TWD in at least a subset of the wholesale models to provide an even more robust view of total pumping and energy requirements**.

4. SSC's CAC on topography

The methodology used by SSC to calculate the value of its topography CAC is exactly the same as the one used in our CAC: that is, to remove BPS from the modelling suite and rely on APH TWD as the only topography driver. The only difference is that we considered a counterfactual scenario where APH TWD is not used at all rather than Ofwat's current modelling suite assuming a 50:50 split between APH TWD and BPS.

While the methodology is the same as our CAC, as part of the general narrative explaining the need for adjustments and the non-appropriateness of BPS in the modelling, we have identified a few points that are worth mention and further study. Some of them are already mentioned in Section 2 above but we have studied in more detail here the issue raised with the number of BPS being unable to reflect the topography of a whole network or the capacity of its sites.

In the example below provided by SSC (Figure 4.1), we clearly see that Ofwat's PR19 and proposed PR24 modelling suite is inadequate as it fails to consider the actual energy or capital maintenance requirements faced by companies. To take an extreme example, if over AMP8 we were to have 466 BPS with a capacity of 30 kW for each of them, Ofwat's current modelling suite would assume them to be as costly as 466 BPS with an associated capacity of 12MW. Based on models relying on BPS as a 'topography' driver, we would therefore receive the exact same allowances in both cases which is completely counterintuitive and inappropriate as it would not be representative at all of our day-to-day operations related to pumping.

One could even say that this creates perverse incentives to build BPS which are much smaller than necessary and to build this strategy based on both the actual costs faced and the likely extra allowance granted by the models, rather than based on actual costs only. APH does not face this issue. Indeed, to the extent possible, using APH creates incentives to reduce the number of BPS and to operate with the smallest number of BPS possible. APH therefore promotes cost-efficiency.

Figure 4.1 Example of the non-appropriateness of BPS as a topography driver

Images above show Hampton Loade's high lift pump hall at 12 MW of installed pumping capacity, and a small distribution booster of 30 kW capacity. Both of these sites would be treated equally in the booster stations per length of main variable, counting as one site each, which is clearly not reflective of the power costs nor the capital maintenance requirements.

Source: [SSC (2023), 'Cost adjustment claim for topography', page 6.]

At the very least, if Ofwat does not want to rely on APH (TWD or total) in all models, it would be more appropriate to consider the *capacity* of BPS, rather than simply counting them which we have shown to be inappropriate and unable to reflect network topography. As a sensitivity, we have tested replacing the number of BPS per length of mains with the capacity of BPS per length of mains.

The alternative driver accounting for the capacity of BPS performs better than BPS per length of mains since, in this case, the associated capacity of BPS is not ignored.

Table 4.1 Statistical performance of the capacity of BPS as a substitute to the number of BPS, WW models

	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Ln(BPS/mains)	0.406**(0.019)	0.407**(0.011)	0.466***(0.006)	0.457***(0.005)	0.341*(0.052)	0.326**(0.049)
Ln(Total capacity of BPS/mains)	0.105***(0.002)	0.104***(0.002)	0.109***(0.001)	0.108***(0.001)	0.120***(0.005)	0.117***(0.006)

Note: *** Statistically significant at the 1% level, ** at the 5% level, * at the 10% level. [1] A positive number means than BPS/mains is less statistically significant than the alternative driver tested, and vice-versa. P-value of the estimated coefficient is provided in brackets.

Source: Anglian analysis from Ofwat's PR24 dataset and the 2022/23 APR.

We note that in TWD models, the alternative drivers accounting for the capacity of BPS also perform better than the number of BPS, with a p-value of 0.000-0.001 as opposed to 0.002-0.006.

While our view is that APH is superior to a cost driver accounting for the capacity of BPS to reflect network topography, the latter would however still be a more suitable alternative to the number of BPS, both in terms of engineering rationale and statistical performance in TWD and wholesale models.

5. SVE's CAC on water network complexity

The methodology proposed by SVE can be divided in two distinct parts: i) use APH WRP in WRP models and total APH in wholesale models as per SES' CAC; ii) use BPS and APH in the same TWD and wholesale models. Since we explain above the merit of using APH WRP, we focus here on the use of APH and BPS in the same models.

First, this proposal faces the same issues highlighted in Section 2. We therefore completely disagree with SVE's position that BPS per length of mains is "the 'true' topography proxy" (page 6 of its CAC). While we agree that 'small boosters are necessary to move water through hilly terrain to serve customers that live in these largely rural areas' (page 6), we have shown in Section 4, through our analysis of SSC's CAC, that simply counting the number of BPS leads to counterintuitive results and does not achieve the desired effect. In any case, the operating and maintenance cost of these smaller stations is smaller than much bigger BPS. Therefore, it is inadequate to ignore the associated size and capacity.

Second, both Ofwat and SVE consider that BPS and APH are both topography drivers, so including both measures in the same model is inappropriate. If BPS were considered as an appropriate cost driver that drives the right incentives (which we heavily contest), it would be inappropriate to build a model relying on these two cost drivers together since they are both supposed to capture network topography. This would run counter to Ofwat's principle to not use variables that intend to capture the same effect. Intuitively, this would be equivalent to using two different deprivation metrics in the same bad debt cost model which no one has considered appropriate.

A similar argument could be made in the wastewater modelling: we find that the joint consideration of the weighted average treatment size (WATS) and the percentage of load treated in STWs above 100,000 people variables in a sewage treatment (SWT) model leads to similar good statistical results, as we can see in Table 5.1 below. However, this is not aligned with the operational rationale and does not mean we should use such a model. The WATS variable and the percentage of load treated in STWs above 100,000 people variable are both supposed to capture economies of scale at large sewage treatment works. Including them both in the same model is counterintuitive as both interact with each other and can hide the true estimated relationship. This is the same with APH and BPS, and the statistical performance cannot be invoked in a context where economic justification is insufficient. In addition to a perverse unobserved interaction between APH and BPS, the good performance of BPS may be partly explained by its correlation with density variables which artificially inflates its statistical performance.

Consequently, we consider that the joint statistical significance of APH (TWD or total) and BPS is spurious as it is not justified from an economic point of view.

Table 5.1 Statistical results of a SWT model with WATS and the percentage of load treated in STWs above 100,000 people

	Model 7
Ln(load)	0.714***
% of load with ammonia consent < 3 mg)	0.006***
% of load in STWs>100k people	0.008***
Ln(WATS)	-0.344***
Constant	-1.51 (0.112)
R-squared	0.91

Note: *** Statistically significant at the 1% level. When the estimated coefficient is not statistically at a level of at least 10%, the p-value is provided in brackets.

Source: Anglian analysis from Ofwat's PR24 dataset and the 2022/23 APR.

To summarise, it is inappropriate to use both BPS and APH in the same model. In addition to the use of BPS being inappropriate, SVE's proposal goes a step further by diluting even more the impact of

APH in the modelling. This moves away from the initial desired effect sought by Ofwat and the industry with the re-inclusion of APH to capture network topography.

II. Reservoirs maintenance

This section reviews United Utilities' (UU) CAC covering the additional costs incurred in maintaining reservoirs. The claim is only partly symmetric, as it relates both to the additional historical costs incurred compared to boreholes and to the increasing expenditures deriving from recent regulatory changes. We focused on the symmetrical component of the claim.

Contents

1. Summary of our position

Overall, we find that the operational and econometric evidence presented by UU is not robust.

We contest the existence of a clear cost trade-off between operating reservoirs and boreholes. Moreover, the company fails to justify the uniqueness of its position beyond the challenges deriving from the new regulatory environment. These do not relate to historical costs and, as such, are not relevant in determining a symmetric cost adjustment.

Looking at our econometric results, the number of reservoirs is a significant variable in only half of the wholesale water models which, as a consequence, are not included in the proposed adjustment. Moreover, the use of alternative drivers, such as the share of water sourced from reservoirs, is not significant and leads to a great degree of volatility in the results. Lastly, we believe the proposed estimation approach is excessively complex, further increasing the fragility and unreliability of the estimates.

As such, **we do not consider the CAC to be valid**.

2. Review of the claim's operational rationale

The claim presents UU's position, which is that (a) some reservoir maintenance costs are partly "outside of management control" and (b) there has been an increasingly challenging regulatory environment. However, the latter concerns mostly future expenditures and, as such, is not relevant in the context of the proposed symmetric adjustment.

UU operates the largest number of dams and reservoirs in the industry, with a relatively high average asset age. However, we disagree with the key points relating UU's unique position with the claim of additional costs incurred and the adjustment proposed.

2.1 There is no singular trade-off between the cost of operating reservoirs compared to boreholes

UU presents the operational challenges related to operating and maintaining reservoirs as opposed to boreholes. In particular, emphasis is placed on the introduction of pumping activity drivers to Ofwat's base cost models which, according to UU, reflect the additional costs to companies who predominantly use groundwater sources. In contrast, UU argues that the WRP cost models do not reflect the extra costs of dam maintenance for those companies which have a higher than average number of reservoir sources compared to groundwater sources, thus leading to unfair outcomes.

However, while it is true that water companies rely on different sources of water, each leading to different costs across the value chain, the relationship between pumping activity, reservoir maintenance and costs is significantly more complex than presented.

First, all **water companies use a multiplicity of sources**, which also include rivers and pumped storage, each leading to different cost levels for pumping, treatment and distribution. Hence, the simple addition of pumping activity as a driver does not imply a clear advantage or disadvantage for a company. Moreover, contrary to UU's narrative, pumping drivers are not included by Ofwat in any of the proposed WRP models, disadvantaging companies sourced predominantly from groundwater, as abstraction pumping costs are not taken into account. Even in the wholesale water models, the inclusion of booster pumping stations per length does not control for the additional costs of sourcing from boreholes, as those pumps are excluded from the BPS variable.

Second, **the cost trade-off between the use of reservoirs and boreholes is not as strong and binary as presented**. UU's claim is that the cost disadvantage for companies carrying out additional pumping activities is offset by savings from lower reservoir maintenance costs. However, if this were true we would not see the significant positive coefficient on pumping variables. Moreover, the correlation between APH_TWD and number of reservoirs, although negative, is weak (c. -0.2). Given the limited correlation, and the presence of alternative water sources (such as rivers and pumped storage), UU fails to justify the need for a parallel adjustment for reservoirs.

2.2 The company fails to fully demonstrate the uniqueness of its position

While it is true that UU operates the largest number of reservoirs in the industry, in relative terms, **UU ranks only third based on the number of reservoirs per property**. This suggests that companies with higher or similar numbers of reservoirs per property (i.e. HDD, WSH or YKY) do not perceive a significant disadvantage from a high reliance on reservoirs.

2.3 Lack of evidence to support that the number of reservoirs as a cost driver

UU argues that Ofwat's objections to reservoir maintenance in previous regulatory periods have now been dismissed, since a specific question on reservoir maintenance has been included in the Base

Cost consultation. In particular, Ofwat's objection concerned the lack of evidence to support the fact that the number of reservoirs is a cost driver.

However, Ofwat's consultation states that "engineering rationale indicates that the number of *high-risk* reservoirs is a more appropriate variable than the capacity of reservoirs" [emphasis added]. **The number of *high risk* reservoirs is a different variable from the one proposed by UU (i.e. their total number)**. Moreover, given the significant change in risk levels following the 2020 Balmforth Report[9], Ofwat's consultation appears to be aimed at controlling for future expenditure increases rather than to adjust for historical cost differences.

3. Review of the claim's econometric justification

Given the questions raised above about the operational support for the adjustment, econometric justification is required.

3.1 Failure of number of reservoirs in the wholesale models

If the trade-off between pumping and reservoir maintenance costs, as argued by UU, were significant, we would expect to identify it at the aggregate level. However, **in the wholesale models the number of reservoirs is significant when used alongside APH_TWD, but not when used with number of booster pumping stations (BPS) per length**. Indeed, we find APH_TWD to be negatively correlated with the number of reservoirs (although weakly, with a correlation of -0.2), while there is a strong positive correlation with BPS per length (0.5).

UU does not propose any adjustment to the wholesale models and excludes them from the calculations. It would be unusual for Ofwat's cost assessment approach to include cost drivers at one level of aggregation and not at another.

3.2 The share of water abstracted from reservoirs is not a significant driver

UU argues that it abstracts the highest proportion of water from impounding reservoir sources. As such, it claims to be disproportionately affected by Ofwat's recognition of power requirements and by the exclusion of dam maintenance requirements.

However, **when replacing the number of impounding reservoirs in UU's proposed WRP models with the percentage of water from impounding reservoirs, the coefficient is largely insignificant** (i.e. it is significant at a 10 percent level in only two models out of six).

In order to examine the impact of different sources on costs, we also tested the impact of separately adding to the WRP models the share of the other water sources (rivers, boreholes and pumped storage). We found that both rivers and pumped storage had strongly significant coefficients (1 percent), respectively negative and positive, while both boreholes and impounding reservoirs are insignificant.

It appears clear that the share of water abstracted from a source can be a significant driver in accounting for additional costs, as it appears to be in the case of pumped storage, but this is *not* true

[9] An independent review of reservoir safety legislation and its application was commissioned by the Secretary of State for Environment, Food and Rural Affairs in 2020. The review was led by Professor David Balmforth. The review report, published in May 2021, provided a comprehensive assessment of the reservoir safety regime in England. It made recommendations for improving the safety regime and strengthening the roles and responsibilities of the regulator, reservoir owners and engineers.

for impounding reservoirs. Furthermore, we believe that no further adjustment to the models is required, as cost differences between the sources of abstraction are already accounted for by pumping and treatment complexity drivers.

UU overlooks the fact that the percentage share of water from a particular source was used by the CMA at PR14 (rivers and reservoirs were included separately in the same model and both were positive but not always significant). However, in its PR19 cost assessment consultation, Ofwat stated, "The source of water may affect pumping and maintenance costs. On average, **we expect impounding reservoirs to have the lowest cost per unit of water**. The other sources are more difficult to rank" [emphasis added].[10]

In its PR19 technical annex, Ofwat subsequently stated that "the water resources plus level of aggregation that we adopted captures the interaction between different services of the value chain and allows for a comparison of costs that internalise inherent choices and trade-offs across the value chain […] Some models were well fitted but lacked an engineering rationale. For example, **we expect costs associated with sourcing water from impounding reservoirs to be relatively low** compared with other sources (e.g. pumped storage reservoirs, rivers and boreholes), however some company models showed the opposite effect" [emphasis added].[11]

Last, the correlation **coefficient between APH_TWD and the percentage of water from reservoirs, despite being negative, is extremely low (-0.03)**, thus further weakening the point made by UU that there is a trade-off between these two factors.

3.3 Excessive complexity and fragility of the proposed adjustment mechanism

UU proposes a combination of econometric and unit cost approaches in estimating the symmetric adjustment.

In general, the CAC appears to be constructed in in an excessively complex fashion, based on three components:

- Part 1: The impact of operating reservoirs vs boreholes
- Part 2: A rise in the number of statutory actions since the 2020 Balmforth Report
- Part 3: A change in the EA flood risk maps requiring additional work to remain compliant with the H&SWA 1974.

Part 1 is the only one which is symmetric and hence relevant for the calculation of the impact on other companies. The impact for other companies is calculated as follows:

- The number of reservoirs is added as an additional regression variable in the six WRP models. The resulting changes in modelled costs (compared to Ofwat's original models) are used to calculate each company's share of the industry change in modelled costs (as a percentage).
- Said percentage is then applied to the backward-looking element of UU's maintenance expenditure. This is calculated as average unit cost, to which catch-up and ongoing efficiency challenges are applied.
- Dam maintenance and avoided power implicit allowances are then subtracted from the derived amount.

[10] Ofwat (2018), "Cost assessment for PR19: a consultation on econometric cost modelling", March, page 16.
[11] Ofwat (2019), "Supplementary technical appendix: Econometric approach", January, page 11.

In particular, **the use of non-econometric models leads to numerous issues**.

Additional implicit allowances need to be calculated in order to avoid double-counting, adding further complexity and uncertainty to the results.

Also, **the average unit-cost approach in calculating UU's historical costs for dam maintenance is not robust**, as it only relies on data from one company, namely UU (despite the addition of both ongoing and catch-up efficiency challenges). Econometric estimates would prove to be more reliable and flexible, especially over time.

Last, the use of a company's percentage of the change in the industry allowance as a scalar of the symmetric adjustment proves to be a further element of fragility, as even **small changes in other companies' allowances can impact another company's allowance**.

3.4 Volatility of results

The proposed models are also prone to high levels of volatility. We tested the use of alternative variables as cost drivers (i.e. replacing the number of reservoirs with the percentage of water from reservoirs and the total volume of water contained in reservoirs), and found significant volatility in the estimates. The volatility was caused both by the change in the estimates, and by how these are passed through the following steps of the calculations.

3.5 Issues with the replicability of the results

We have encountered issues in replicating the estimates presented in the claim, obtaining markedly different values when following the methodology as described.

In particular, we noticed a difference in the number of observations in UU's results, which could be due to part of the data being excluded or dropped during UU's analysis. These details are not explained in the CAC. It is, however, telling that the results can vary markedly even when attempting to replicate the same approach.

Moreover, the construction of the reservoir variable to deal with zero values appears to be an approximation and may need refinement in terms of data quality.

III. Metering

This section covers companies' responses to the base cost consultation on the exclusion of metering as a cost driver in the retail modelling suite, as well as the two cost adjustment claims (CACs) that have been submitted on the topic of increased meter renewal and replacement costs by South East Water (SEW) and Southern (SRN), respectively.

Contents

1. Summary of our position

Metering penetration remains an important and obvious cost driver from an operational perspective, for both us and the broader industry (as discussed below). Higher costs appear in both retail (for example, from meter reading and because metered customers are more likely to query their bills) and wholesale (due to metering renewal and replacement). Ofwat historically recognised these costs in its cost allowances in order not to disincentivise companies from pursuing metering strategies.

We note with concern that the exclusion of the proportion of metered customers from Ofwat's retail models would lead to insufficient (or potentially zero) allowance for a key cost driver. We acknowledge the continued poor statistical performance of the cost driver in the retail models (in particular the bottom-up, Other Cost models). The inability of the bottom-up retail models to detect a common-sense cost driver is another illustration of why these models can no longer reasonably form a reliable basis for the retail cost assessment at PR24 (discussed below).

In light of this, we recommend that:

1. Ofwat still considers meter penetration as a cost driver in the subset of retail cost models where it performs reasonably well (with the expected sign, magnitude and acceptable p-value) – which, at this point in time, is the top-down model RTC1.
2. Ofwat only uses the subset of top-down, total cost models for actual revenue determinations (whilst bottom-up models may still be considered for cost driver validation, where appropriate).

Outside of the scope of the costs covered by the retail models, we reiterate that increased metering penetration also increases metering renewal and replacement costs for companies with higher levels of metering penetration. These differential renewal costs are not explicitly provided for in the broader base cost modelling suite, and adversely disadvantages companies with higher levels of metering penetration. We thus support the symmetrical cost adjustment claims (CACs) made by SEW and SRN on this basis.

2. Responses to the base cost consultation (retail)

In the May base cost modelling consultation, we emphasised the continued strong operational rationale for metering as a cost driver. We highlighted the following reasons:

- Continued meter reading costs
- The increased customer contacts from metered customers, and
- Meters are relatively short-lived assets that require frequent and costly replacement.

Given the extent of variation in meter penetration levels across companies (ranging from 48 percent to 90 percent in 2022/23), we would thus expect differential metering penetration to have a continued meaningful on company costs. A positive-signed metering cost driver in the retail models also has the policy benefit of incentivising continued metering rollout and rewarding companies with higher metering penetration rates.

We note that the majority of the industry shares our position, with eight of the 12 other companies that took a view on metering also disagreeing with Ofwat's proposal to remove the cost driver from the retail models. These companies all cited the strong operational rationale for retaining meter penetration as a cost driver, reiterating different variations of the same arguments that we had made.

We also note that the independent *Retail services efficiency review* by PwC for Ofwat highlights: (i) the increased customer service costs resulting from increased metering penetration;[12] and (ii) the broader benefits of increased metering, such as earlier leakage detection and demand reduction.[13] These points are also raised by other companies in their responses to the base consultation.[14] This evidence supports the incentive rationale for maintaining metering as a cost driver, where it is of the appropriate sign and performs sufficiently well from a statistical perspective.

In contrast, from the minority of companies that agreed with Ofwat that metering penetration should be excluded as a cost driver, the central reason cited is the cost driver's poor statistical performance.

We acknowledge that metering penetration does not perform well statistically, especially in the Other Cost models. In our May response to the base cost consultation, we suggested that Ofwat should still consider the cost driver in the subset of models where it performs adequately (at this point in time, only the total cost model RTC1) and reconsider the variable's broader inclusion as more data becomes available (post the potentially distortionary impact of COVID-19).[15]

More broadly, we are concerned by the Other Cost models' poor performance and lack of coherence. Their inability to detect the impact of metering penetration on costs may well be due to the miscellaneous nature of the 'other' cost category and the resulting challenges in identifying a coherent and consistent set of cost drivers.

---

[12] PwC (2022), 'Retail services efficiency review 2022. Report for Ofwat.' December, page 48.
[13] PwC (2022), 'Retail services efficiency review 2022. Report for Ofwat.' December, pp. 48-50.
[14] See SES and South East Water's (SEW) consultation responses, respectively. For example, SES notes that metered consumers consume 15% less per capita than unmetered customers.
[15] As noted in our consultation response in May, metering is still of the right sign, expected magnitude and acceptable levels of statistical significance in the total retail cost model RTC1 (where at the time it had a p-value of 0.17 – in line with the level accepted by Ofwat on other modelling areas). These same results hold when the consultation models are updated with 2023 company APR data (not shown here).

The Other Cost models' inability to detect a common-sense cost driver is another illustration of why we believe that the bottom-up models can no longer reasonably form a reliable basis for the retail cost assessment at PR24. We suggest that Ofwat bases its retail cost determinations solely on the more reliable set of top-down models (with bottom-up models used at most for cost-driver verification for the appropriate subset of costs), for the following reasons:

- **The poor performance and lack of coherence of the Other Cost models**. The Other Cost models have comparatively little explanatory power, with an average adjusted R-squared from Ofwat's consultation models of only c. 0.12 (compared to c. 0.66 across the total cost models – see table 2.1 below). Removing metering would also leave at most two cost drivers in the miscellaneous Other Cost category.
- **The collective poor performance of the bottom-up models relative to the top-down models**. The bad debt models' performance has also deteriorated, with an average adjusted R-squared some 0.12 lower than at PR19 (based on the average across consultation models, updated with 2023 APR data). As illustrated in the table below, Ofwat's base cost consultation models indicate that the 'triangulated' adjusted R-squared of the bottom-up models is much worse than the top-down modelling at PR24.

Table 2.1 Bottom-up vs top-down model performance at PR19 and PR24

<table>
<tr><th rowspan="2"></th><th colspan="2">Adj R squared</th><th rowspan="2">Weight at PR19</th></tr>
<tr><th>PR19</th><th>PR24 (consultation)</th></tr>
<tr><td>Bad Debt (bottom up)</td><td>0.78</td><td>0.66</td><td rowspan="2">25%</td></tr>
<tr><td>Other Costs (bottom up)</td><td>0.14</td><td>0.12</td></tr>
<tr><td>Total Costs (top down)</td><td>0.69</td><td>0.66</td><td>75%</td></tr>
</table>

Note: Average adjusted R-squared across the respective sets of bad debt, other cost and total retail cost models. PR24 adjusted R-squared values from simple average across base cost consultation models, updated with 2023 APR data.

Sources: (i) Anglian update to retail models in Ofwat (2023), ' Econometric base cost models for PR24', April, (ii) Ofwat (2019), 'PR19 Final determinations. Securing cost efficiency technical appendix.', page 173.

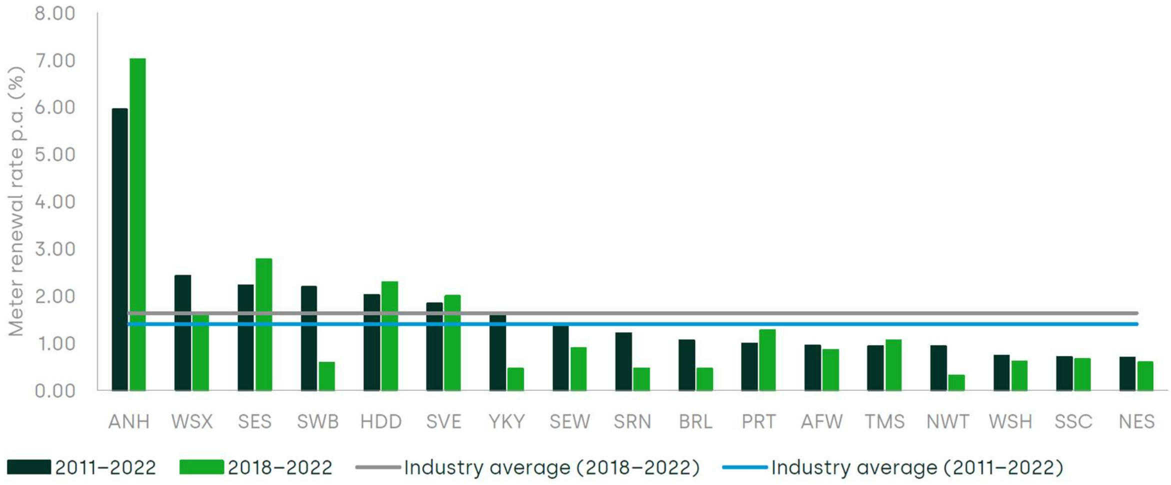
3. CACs on metering renewal/replacement

We note that SEW and SRN have submitted CACs for above-average meter renewal/replacement activity expected over the coming price control period.[16] Both state that the adjustment is required because meter renewal/replacement activity is not explicitly modelled as a cost driver in the base cost modelling suite (in this case relating to Water Network Plus costs). We also note that Ofwat considers that these CACs would be symmetrical (at least partly).[17]

We support these symmetrical CACs, given the relatively short asset lives and costly replacement of meters (as highlighted previously) and the fact that these costs disproportionately affect companies with higher levels of metering penetration. The need for this cost adjustment is augmented by the broader importance and public benefit of incentivising metering rollout, and thus providing sufficient allowance for the expanded stock of meters' eventual replacement. Illustratively, as shown in SEW's CAC, Anglian has consistently had one of the highest metering penetration rates across the industry – and thus historical renewal rates significantly greater than the industry average (illustrated in Figure 3.1 reproduced below).

[16] As outlined in Oxera (2023), 'A review of cost adjustment claims for PR24. Prepared for South East Water', 9 June, section 2 and SRN (2023), 'Meter replacement cost adjustment claim', 9 June.
[17] Ofwat (2023), 'Cost adjustment claim summary – August 2023.'

Figure 3.1 Average meter renewals (2011-22 and 2018-22)



Source: Oxera (2023), 'A review of cost adjustment claims for PR24. Prepared for South East Water', 9 June, p. 7.

Whilst we have not attempted to assess and replicate the CAC calculations submitted by SEW and SRN, we consider their general approaches to calculating the value of the claim to be intuitive, reasonable and consistent with Ofwat's CAC guidance. More specifically,

- As suggested by both companies, it seems reasonable that the implicit allowance should be determined in relation to the industry average annual meter renewal/replacement activity over the historical period considered in the base cost models. As suggested by SEW, the implicit allowance can further be adjusted for the extent that renewal activity is correlated with (and thus implicitly funded by) other cost drivers in the Water Network Plus models.
- We also agree with the high level approach suggested by both companies to base the (net) value of the claim on the basis of an efficient meter replacement unit cost (calculated from historical industry data), multiplied by the extent that the company's expected meter renewal/replacement activity is greater/less than what is implicitly funded by the base cost models.

IV. Economies of scale at water recycling centres

This section covers companies' responses to the base cost consultation on the best way to capture economies of scale at water recycling centres (WRCs) and to provide an update of the statistical performance of the models in light of the 2022/23 data.

We welcome Ofwat's attempt to develop a new measure studying large WRCs at a more disaggregated level and note that, with the exception of Thames (TMS), the weighted average treatment size (WATS) variable gains unanimous support within the industry. The only argument mentioned by TMS is the difference in the measurement of bands 1-5 compared to above band 5 where the variable captures the contribution of each large STWs individually. We do not consider it to be an issue since the economic rationale behind the construction of this variable is to be able to distinguish the continuous decrease in unit cost for larger STWs. There is then no doubt about its utility at PR24. The outstanding question is whether it is appropriate to use only the WATS variable or to complement it with one alternative driver, e.g. the percentage of load treated in bands 1-3 (used at PR19, thereafter pctload13) or the percentage of load treated in WRCs higher than 100k people (a refined version of the PR19 driver, thereafter pctload>100k).

We study this question in two steps. We first focus on the WATS versus the pctload>100k and second on the WATS versus the pctload13.

Contents

1. Summary of our position

The evidence in favour of the WATS variable (or against the pctload13 and the pctload>100k) is even stronger than at the consultation phase. The clear superiority of the WATS variable has been strengthened from a statistical perspective with the inclusion of the 2022/23 data, particularly when compared to pctload13 which has very little explanatory power for treatment costs (p-value of 0.26 in the Treatment model versus 0.003 for the WATS variable).

We therefore reiterate our May base cost consultation response that models with WATS are clearly superior to any model that includes a population threshold variable (representing a step change in unit costs either side of this threshold) as these alternative variables cannot capture the *continuous* fall in unit costs implied by an increase in the size of large WRCs. As indicated in our CAC, in the event of the WATS variable being the only cost driver used to capture economies of scale, we would not need any cost adjustments claim on this topic as our operating circumstances would be directly reflected in the modelling.

As mentioned below, if a cost driver is unable to explain the continuous decreasing costs it is supposed to (here treatment costs) it should not be use in top-down models using a wider cost aggregation (here total Water Recycling Network Plus costs) since any statistically significance of the estimated relationship would be spurious.

Regardless of the choice of the most appropriate cost driver, the existence of economies of scale at WRC works is clear and gains unanimous support across the industry. Therefore, it is not appropriate to rely on Water Recycling Network Plus models that do not include any measure of economies of scale (Water Recycling Network Plus models 10 and 14), as the only effect of such models is to dilute the impact of economies of scale which is counterintuitive and not aligned with the industry's view or Ofwat's attempt to develop alternative measures of economies of scale. It is clear that these top-down models need to be dismissed.

2. WATS versus the percentage of load treated in WRCs higher than 100k people

As mentioned in our May base cost consultation response, WATS represents a clearly superior alternative to capture economies of scale to a population threshold-based variable such as pctload>100k. This is, for example, demonstrated by our illustration of using alternative models derived with different population thresholds such as 150k, 200k or 250k, each of them showing different results (see page 2 of the appendix to our base cost consultation response). None of them are particularly more suitable, although the higher the population threshold the better the ability of the model to capture economies of scale at large WRCs. However, the higher the population threshold the lower the number of observations available in the regression analysis. There is, therefore, a trade-off between both effects which is handled particularly well by the WATS variable since it allows a *continuous* relationship between costs and the size of WRCs without *arbitrarily* imposing a tipping point beyond which economies of scale start to occur.

As mentioned in the introduction, nine out of the ten WaSCs[18] highlight the merits of the WATS variable and its ability to capture economies of scale at large WRCs. The only company arguing against its use, TMS, mentioned a presumed inconsistency in measuring the variable but we disagree with this point. Indeed, the aim of the WATS variable is to capture the lower unit costs of larger WRCs which is made possible by an analysis led at a disaggregated level. If more disaggregated data were to be available for bands 1-5, the actual results of increased economies of scale as the size of STWs increases would remain unchanged.

The economic intuition indicating that WATS is more suitable than pctload>100k is confirmed by the respective performance of these two cost drivers in the models proposed by Ofwat. While there is not any material difference in statistical significance of the cost drivers in Treatment models, there is a substantial gap in Water Recycling Network Plus models, with pctload>100k not even being significant in the Water Recycling Network Plus models (with p-values of 0.207 and 0.115).

[18] Treating SVE and HDD as a single unit, SVH

Table 2 Statistical performance of WATS and pctload>100k in Treatment and Water Recycling Network Plus models

	SWT2	SWT3	WWWNP3	WWWNP4	WWWNP7	WWWNP8
Ln(WATS)	NA	-0.240***	NA	-0.091***	NA	-0.097***
pctload>100k	-0.008***	NA	-0.002(0.207)	NA	-0.003(0.115)	NA
Difference in p-values[1]	0.003		0.198		0.113	

Note: *** Statistically significant at the 1% level. [1] A positive number means that WATS is more statistically significant. This aims to help illustrating the point of the higher statistical significance of the WATS variable, but the difference is displayed only for purposes of qualitative comparison.

Source: Anglian's analysis based on Ofwat's PR24 dataset and the 2022/23 APR.

Given the strong economic rationale of the WATS variable, its clear superior statistical performance compared to the pctload>100k, and the industry consensus on this question, there is no valid reason to use pctload>100k to capture economies of scale.

3. WATS versus the load treated in bands 1-3

Compared to the consultation phase, we now have even more concerns on the inability of the pctload13 to accurately reflect the higher unit costs of smaller size bands.

Conceptually, this variable faces the same issues as the pctload>100k since it assumes an *arbitrary* threshold, bands 1-3, beyond which diseconomies of scale no longer occur. Moreover, the statistical performance of this variable in the sewage treatment model is even more concerning and clearly does not meet Ofwat's standards.

The p-value is particularly high, at 0.26, which confirms the weaknesses faced by this variable.

Table 3 Statistical performance of WATS and pctload13 in Treatment and Water Recycling Network Plus models

	SWT1	SWT3	WWWNP2	WWWNP4	WWWNP6	WWWNP8
Ln(WATS)	NA	-0.240***	NA	-0.091***	NA	-0.097***
Pctload13	0.026(0.258)	NA	0.022**	NA	0.023**	NA
Difference in p-values[1]	0.258		0.022		0.111	

Note: *** Statistically significant at the 1% level, ** at the 5% level. [1] A positive number means that WATS is more statistically significant. This aims to help illustrating the point of the higher statistical significance of the WATS variable, but the difference is displayed only for purposes of qualitative comparison.

Source: Anglian's analysis based on Ofwat's PR24 dataset and the 2022/23 APR.

While the statistical performance of pctload13 is better in Water Recycling Network Plus models and sufficiently high, this is only because it acts as a proxy for density with a correlation of about 0.70 with both WAD measures .[19] [20] If we replace the pctload13 variable by any of the two WAD measures, we find that the estimated coefficient of the WAD measure is always statistically

[19] This is also why we have proposed a different alternative in our May base cost consultation response by giving weight to WAD LAD from MSOA as a substitute to pctload13. This could be a second best option as a substitute to pctload13 (and as a complement to WATS) if Ofwat thinks there is still a need to use a second model that captures higher unit costs implied by sparse areas/smaller WRCs.

[20] Based on the logarithm of WAD measures.

significant at the 1 percent level and the R^2 improves (i.e. the modelling improves). The statistical performance of pctload13 in WWWNP models is therefore spurious and coincidental.

If a variable specifically constructed to explain treatment costs is statistically insignificant in a treatment model, it should be completely removed from the modelling suite as it highlights its inability to capture the desired effect, i.e. the lack of economies of scale in smaller WRCs. Using a variable in top-down models which is not being used in bottom-up models would depart from Ofwat's practices at PR19 (and PR24 to date) and be against the economic rationale as pctload13 does not directly explain collection costs.

V. Growth at water recycling centres (WRCs)

This section provides our view on the best way to determine allowances for costs related to growth at water recycling centres (henceforth, WRC growth costs) at PR24. In our assessment, we have also considered the two cost adjustment claims (CACs) submitted by Severn Trent (SVE) and Wessex (WSX) on this topic.

Contents

1. Summary of our position

We propose that WRC growth costs be determined by reincluding them within the base cost models, as per the approach at PR19. When including these costs in the different dependent variables, the performance of these models is high enough to be relied upon. We believe this is the most robust approach at this stage as all the different attempts to build an accurate and robust standalone model have failed to date.

The main reason we have come around to support the inclusion of these costs within the base cost models is because we have not seen evidence that a robust allowance can be determined for these costs under any other method. We remain open to an alternative model to determine allowances but believe that the base cost models should be used where no sufficiently robust alternative exists.

However, since the consideration of WRC growth costs in Botex Plus is not perfect either, we recommend companies' business plans are assessed thoroughly in this respect and an uplift to the implicit base cost allowance considered if there is sufficient evidence that the planned AMP8 expenditure is efficient.

2. Outcomes when including these costs within the base cost models

When including WRC growth costs in the base cost models, companies' resulting implicit allowances are within a reasonable range of their actual expenditure and thus we consider that this is a reasonable starting point for determining future allowances.[21] The ratio of these companies' actual expenditure to their implicit allowance ranges from 49 percent to 158 percent, similar to the range of efficiency scores estimated from the base cost models as a whole.

This can be contrasted with the modelled costs under the Arup model where there is significantly more divergence between actual expenditures and the modelled costs: the ratio of the same companies' actual expenditure to their predicted expenditure ranges from 69 percent to 414 percent. Moreover, whereas the base cost models yield a total implicit allowance for the industry of £427 million compared to total actual expenditure of £522 million, the Arup specification models a

[21] Thames Water and Yorkshire Water are two clear outliers, driven by their extreme actual expenditures.

total industry expenditure of £229 million, which would leave the industry significantly underfunded for WRCs growth costs.

The main reason for our preference to use this method to estimate allowances, however, is that no robust alternative has been developed. We discuss the issues with these alternatives in the section below.

3. Alternative models for estimating WRC growth costs

Ofwat commissioned Arup to develop an alternative econometric model to determine allowances for companies' WRCs growth costs. Arup proposed four models with the following specifications.

Figure 3.1 Arup's proposed WRC growth costs econometric models

	10-year capex cumulative (2011-12 to 2020-21)			
Variable name	1	2	3	4
PE change served by WwTWs (000s)	0.00150	0.00162	0.00143	0.00144
Load receiving tertiary treatment (%)		0.0183	0.0162	0.0150
Volume WW change (Ml/yr)			0.00000153	0.00000187
Load treated in WwTW size bands 1-3 (%)				0.0315
Constant	3.623	2.525	2.484	2.373
Dependent variable	GWwTW capex (ln)	GWwTW capex (ln)	GWwTW capex (ln)	GWwTW capex (ln)
Estimation method	OLS	OLS	OLS	OLS
N	10	10	10	10
R^2	66%	77%	80%	81%
RESET test	Pass	Pass	Pass	Pass
VIF score (mean)	1.0	1.0	1.3	1.4

Note: Arup's preferred model is Model 2. Coefficients in dark red are significant at the 1% level; those in light red are significant at the 10% level.

Source: Arup (2022), 'Assessment of growth-related costs at PR24', May, page 42.

Arup's preferred model was Model 2, as this yielded a sufficiently high R-squared value and all of the coefficients were significant at the 10 percent level or better. We focus on this model throughout this discussion, as the other models were even less robust. While at face value this model is sufficiently statistically robust, there are a number of issues with determining allowances under this approach:

- As cited by SVE in their cost adjustment claim and by Arup themselves in their Ofwat report[22], **no viable cost driver could be found to represent the capacity headroom of the wastewater treatment network**. As a result, the main driver behind investment to expand capacity at WRCs is not included in the models.
- Both WSX and SVE note in their cost adjustment claims that **other important drivers of WRCs growth costs - tightened environmental discharge permits - are not included in the**

[22] Arup (2022), 'Ofwat: Assessment of growth-related costs at PR24: Final Report', May, p.39; Severn Trent (2023), 'PR24 Cost Adjustment Claims', June, p.27.

Arup model .[23] These include the dry weather flow (DWF) and flow to full treatment (FFT) permits, which regulate the amount of flow that a WRC must be able to treat, but also the knock-on tightening of environmental consents which come with this. The Arup model does not capture either of these factors.

- **The Arup model often provides a positive allowance to companies even where they are expecting to have negative population growth over the next AMP**. This can be seen in the chart below, showing the relationship of final allowances to projected PE growth over AMP8, where Northumbrian Water and Yorkshire Water, for example, both receive positive allowances despite having significantly negative growth forecasts.

Figure 3.2 Relationship of Arup modelled AMP8 allowances to forecast change in population equivalent served over AMP8

Note: Arup's preferred model is model 2 of the models outlined in Figure 3.1 above.

- Arup's proposed solution to the lumpy nature of WRC growth costs was to take the 10-year totals for expenditure and the different cost drivers (this equated to 12-year totals in our updated models, using data up to 2022/23). **Arup's approach has the effect of reducing the number of observations to 10 (one sum for each company), and thus reduces the robustness of the models significantly**. This was also noted by SVE in its cost adjustment claim.[24] As a potential solution, we have looked at taking five-year rolling totals (or differences, for variables such as population equivalent and volume) of companies' expenditure and cost driver data. These models actually perform worse than the 10-year or 12-year models, as many companies have negative population growth over five-year aggregations, but still spent significant amounts on growth at WRCs. The poor performance of these models, even with these different specifications, further undermines their use.
- One of the weaknesses of these separate econometric models is that they cannot produce reasonable predictions of historical costs, and this **produces a very wide range of historical company efficiency scores**, which would usually be used to set a catch-up efficiency challenge as in the base cost models. As mentioned above, whereas the base cost model efficiency scores range between 49 percent and 158 percent when excluding the two outliers YKY and TMS, the Arup models produce efficiency scores ranging much more widely (when calculated over the last five years ranging between 69 percent and 414 percent, when excluding the same two companies). This further indicates that they are a poor predictor of company expenditure.
- We have also looked at removing the best and worst performing companies in the Arup models, which are Yorkshire and Thames Water respectively. **Removing these outlier**

[23] Wessex Water (2023), 'WSX09 – Annexes – Base cost adjustment claims', June, pagev4; Severn Trent (2023), 'PR24 Cost Adjustment Claims', June, page 22.
[24] Severn Trent (2023), 'PR24 Cost Adjustment Claims', June, page 27.

companies does not lead to significant improvements in the models and no improvement in the significance of the cost drivers, other than growth in population equivalent served.

VI. Combined Sewers and Urban Rainfall

This section provides a review of Yorkshire Water's (YKY[25]) and United Utilities' (UU's[26]) symmetric CACs concerning the introduction of a combined sewers (CS) driver in the wastewater econometric models. In addition, it provides further evidence against the use of urban rainfall (UR), a measure proposed in Ofwat's consultation and closely linked to CS.

Contents

1. Summary of our position at base cost consultation

In our response to the base costs consultation[27], we strongly disagreed with the inclusion of urban rainfall (UR) in the Collection and Water Recycling Network Plus models. The measure suffers from a number of issues, both from an operational and an econometric perspective.

First, rainfall is not homogeneously distributed over a company's area. As such, **Ofwat's measure does not represent the actual urban rainfall in a company's region**, but rather the overall average regional rainfall, weighted by the company's degree of urbanity. The gap between the UR metric and the actual phenomenon is particularly relevant for companies serving larger geographic areas, which are hence likely to be less homogeneous both in terms of rainfall levels and urbanisation. In this regard, we serve the largest region of any WaSC. More precise GIS data should be used, so as to ensure that only the amount of rain falling within urban (catchment) areas is taken into account, while higher-than-average rainfall in non-urban areas is disregarded.

Second, **total annual rainfall is not the driver of costs, rather it is the *intensity* of the rainfall that determines both the capex requirements (e.g. in terms of mains diameter) and the level of opex (e.g. pumping costs)**. Indeed, there is limited correlation between a company's urban rainfall and sewage expenditure or faults. The correlation coefficient between UR per length and the number of internal sewer flooding incidents per length is only equal to 0.12, while the correlation with the

[25] Yorkshire Water (2023), "Combined sewers CAC", June, available here.
[26] United Utilities (2023), "CAC_002 – Drainage Cost Adjustment Claim", June, available here.
[27] Anglian Water (2023), "ANH PE24 response template v0.1", May, available here.

expenditure to reduce risk of flooding in sewage collection per length is 0.04[28]. A superior measure could be based on the number of extreme rainfall events in a year - for example, setting the threshold at storms with a rainfall intensity greater than 25mm/hr.

Third, **UR is constructed starting from urbanity, which is a close proxy of population density, which is already captured in the collection models**. The correlation coefficient between the percentage of urbanity of a WaSC and the various measures of density ranges between 0.81 and 0.87. Ofwat's proposed approach adds the log of urban area multiplied by the rainfall, which is equivalent to adding two variables into the model specification with the same coefficient, namely urban area and rainfall. As a consequence, when switching from total annual rainfall to urban rainfall, we observe a larger decrease in the density coefficient. These observations are particularly relevant as, across companies, the variability in terms of urbanity is significantly higher than that in terms of rainfall (see the table below).

Table 1.1 Variability of rainfall and urbanity

	Urban area %	**Avg annual rainfall (mm)**
ANH	15.7%	631.3
WSH	12.9%	1444.4
NES	16.5%	925.4
SVH	24.6%	781.1
SWB	7.9%	1294.5
SRN	28.3%	837.0
TMS	39.4%	735.2
UU	30.4%	1369.2
WSX	18.4%	913.6
YKY	25.2%	896.2
Ratio max/min	5.0	2.3
S.D. / average	41%	27%

Source: Anglian Water, based on Ofwat data (available here).

Fourth, we noted how **some of the issues generally linked to rainfall, such as water infiltration, are caused by other drivers such as soil type, sewer condition and height of water table**. It is, for example, widely accepted that ground water enters sewers via infiltration and in some catchments these volumes exceed water ingress from rain. Soil types can lead to higher infiltration as a result of their tendency to move and cause sewer disjointing.

In addition to these points, we noted that, while not directly relevant to the wastewater modelling, **low level of rainfall is detrimental to companies in terms of water supply as it implies higher costs**.

In addition to these issues regarding the use of UR, below we examine the use of CS and UR.

[28] Anglian Water, based on data from Ofwat (2023), "PR24 Cost Assessment Master Dataset, Wholesale Wastewater Base Costs v4", April, available here. Data on the number of ISFs is taken from the APR data tables for the period 2017-22.

2. Proposal of a combined sewers (CS) driver

The industry was divided in its response to the inclusion of a UR driver in the wholesale regression models. Four companies opposed it (ANH, NES, SRN, SWB) while six supported it (WSH, SVE, TMS, UU, WSX, YKY). In addition, two of the companies in favour of UR (YKY and UU) have proposed the inclusion of a variable controlling for the share of CS in the Collection and Network Plus models.

The rationale supporting both claims is that the presence of CS leads to increased flooding events (ISF) when there is heavy rainfall, and thus to additional base costs and a worsening in performance. Since the issue cannot be resolved without large upfront investments in converting CS, the companies presented the claims in order to cover these additional base costs.

Given the close relationship between UR and CS, both the claims support the simultaneous use of the two variables. In particular:

- YKY proposes to include the percentage of CS as an additional independent variable, alongside UR;
- UU proposes to create an interaction term (i.e. the product of CS and UR).

3. Critique 1: Introducing CS does not address existing issues of UR

Despite the reasonable operational justification provided in support of combined sewers, the proposal is plagued by the same structural issues affecting UR, in particular with regards to data granularity.

Combined sewers collect both rain and foul waste water and, as such, are more susceptible to increases in the amount of water transported, which can translate into ISF events. These dynamics require the *simultaneous* persistence of three main factors:

i. Combined sewers
ii. Abundant (and especially intense) input of water in the sewers, which can largely be traced to rainfall events
iii. Urban and densely populated areas.

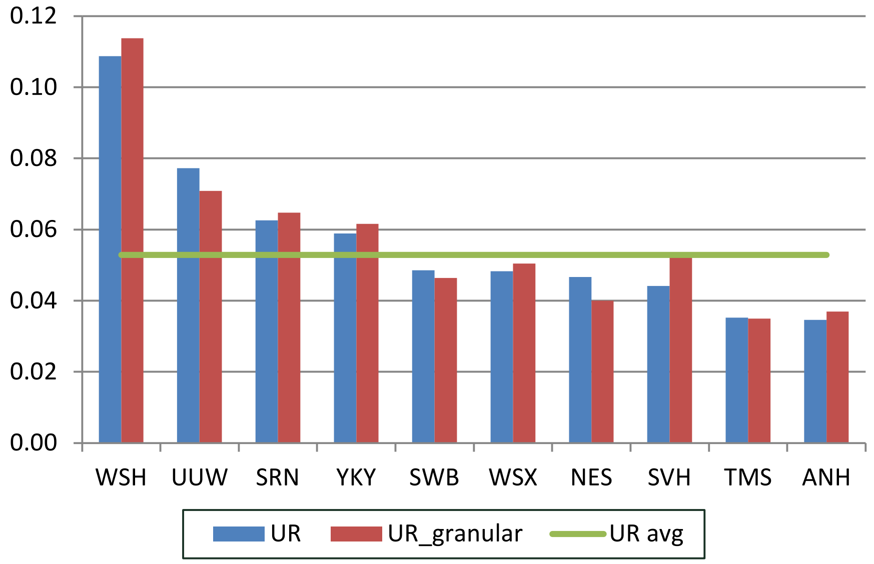
In order to ensure the geographical simultaneous persistence of these three factors we need geographically granular data for CS as well, as already pointed out for urbanity and rainfall in the context of our critique of UR.

To further prove this point, we have constructed a "granular" version of UR, based on MSOA-level data (as opposed to Ofwat's company-level figures).[29] This new specification of UR ensures that rainfall levels are only considered if they are recorded in urban MSOAs, thus improving the precision of the metric.

In Figure 2.1 below, we show how UU's gap from the industry average (over the years 2012-22) is reduced from 46 percent to 31 percent by this refined analysis.

[29] The rainfall data is produced by MetOffice and subsequently mapped to MSOA areas using: ONS (2023), "MSOA (Dec 2011) Boundaries Generalised Clipped (BGC) EW V3", available here.

Figure 2.1 Standard vs "granular" Urban Rainfall



Source: Anglian Water, based on Ofwat and MetOffice data.

Although the new measure still points to marked differences across companies, it demonstrates that data precision can have a significant impact on output. It also illustrates the potential importance of geographically granular data on CS, which at the moment is not available across companies.

Besides granularity, while the use of CS is not directly impacted by the other issues concerning the use of UR,[30] these remain relevant, as the use of CS is advocated in combination with UR, as the two metrics are presented as very closely linked.

4. Critique 2: lack of cost evidence

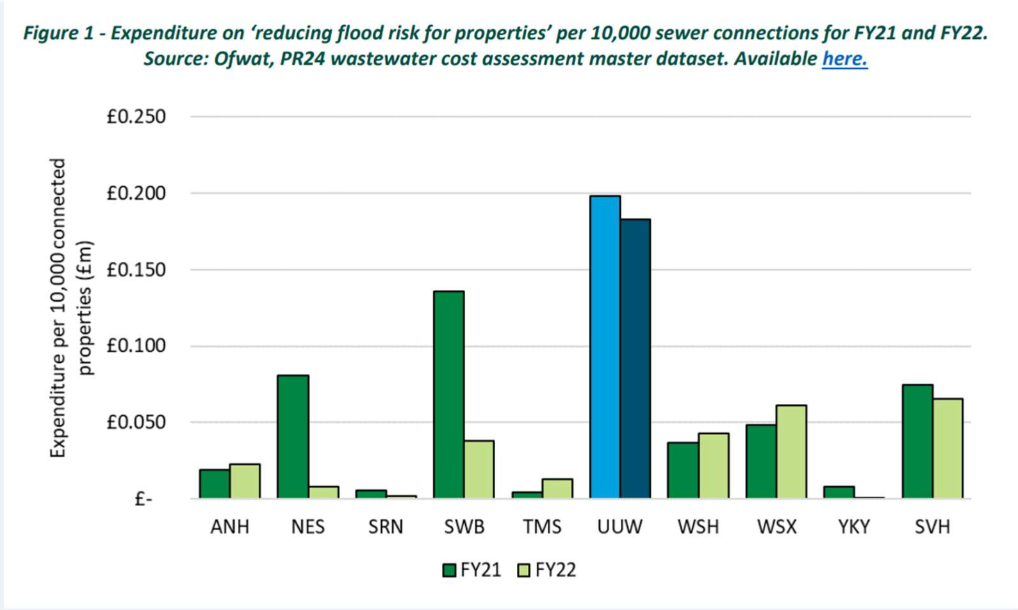
Ofwat requires companies submitting CACs to provide clear evidence of a material link between the claim and the expenditures incurred.

UU demonstrates that it spent more than other companies on managing flood risk ("reducing flood risk for properties") in the first two years of AMP7. However, within this measure YKY spent the least, as is shown in the figure below (taken from UU's CAC). If no other area of expenditure is identified as higher as a consequence of combined sewers, then this strongly undermines the validity of YKY's claim.

More generally, no other cost item is actually identified, as the companies tend to vaguely refer to "increased expenditure on incident response".

[30] That is, the need of a measure of rainfall intensity, the correlation with population density, and the issue of water infiltration.

Figure 2.1 Expenditure on reducing flood risk for properties



Source: United Utilities (2023), "UUW_CAC_002", June, page 7, available here.

The addition of data from 2022/23 only makes the above analysis even more stark.

5. Critique 3: Superiority of CS over UR (and over a combination of the two)

Going back to the list of required factors increasing the risk of ISF presented above, namely:

i. Combined sewers
ii. Abundant and intense rainfall events
iii. Urban and densely populated areas, reflecting both the amount of rainfall collected (i.e. catchment area) and the amount of wastewater produced,

we note that i. represents the necessary condition for factors ii. and iii. to interact. In the absence of i., the flows of rain and wastewater would remain separate.

As such, given i., the intensity of rainfall represents a co-determinant of flooding, alongside the amount of wastewater produced. In this regard, a measure of property density (e.g. property/km of sewer) could control for the impact of the amount of wastewater produced on the number and cost of ISFs more effectively than UR.

It thus seems more appropriate to use a measure (CS) that indicates the percentage of the network exposed to a certain risk – rather than combining it with an imprecise indicator of the risk itself. Combining CS and UR (as proposed by UU) without the use of granular data indicated above leads to the creation of an unreliable composite variable.

Moreover, using the two variables in parallel (as proposed by YKY) reduces their significance level, while leaving unresolved the issues already mentioned regarding UR. Despite the VIF statistic being below the threshold suggested by Ofwat, YKY admits the existence of correlation between the two variables (although "noisy and imprecise").

Last, as already mentioned above, CS is not affected by the other issues concerning UR, especially as it is uncorrelated with the density measures. As such, the use of CS alone appears to represent a superior alternative to any scenario including UR.

6. Critique 4: econometric results

In further support of Critique 3, we ran the proposed wholesale econometric models using only CS as an additional cost driver in place of UR. We note that in this case the R-squared is immaterially lower than when UR and CS are combined (at most a 0.4 percent decline), as shown in the table below.

Thus, CS appears to have a greater explanatory power than UR, whose addition has also only a limited impact.

Table 2.1 R^2 across various model specifications

	UR only (Ofwat)	UR*CS (UU)	UR+CS (YKY)	CS only
Average R^2	0.958	0.967	0.965	0.964

Source: Anglian Water, based on Ofwat models.

VII. Density

This section covers companies' responses to the base cost consultation on the two questions related to density and to reiterate our view on the best way to capture density in the modelling across the three wholesale service areas.

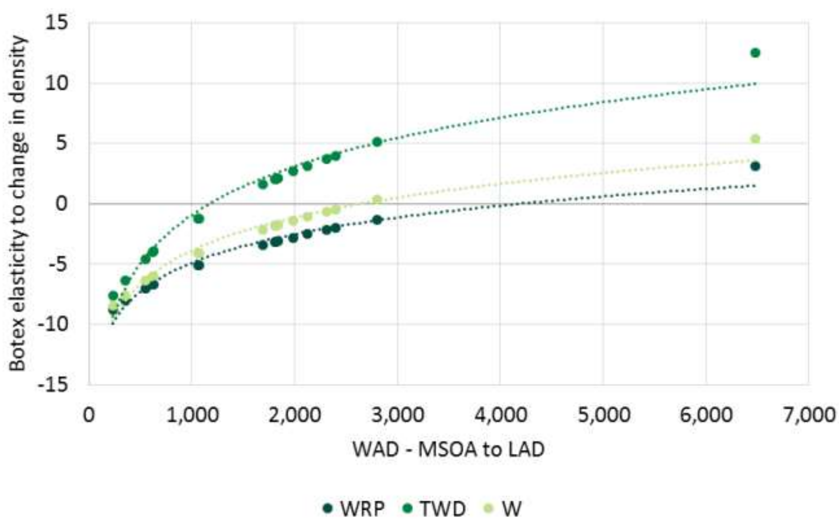
1. Summary of our position

In our May base cost consultation response, we expressed our preference for the refined version of the PR19 density driver, WAD LAD from MSOA. This is because we think Ofwat has made the necessary adjustments to the variable by fixing the different inconsistencies arising from the mapping between LADs and companies' operating areas over the last couple of years. Indeed, WAD LAD from MSOA is derived the other way around compared to the PR19 version, i.e. first computing the specific population density faced by the company and second putting it in perspective with the LAD it serves. We think it is the correct way to proceed. While representing an improvement of the PR19 LAD measure, it also improves the consistency over time in the way density is modelled.

We have expressed strong concerns about the WAD MSOA variable which produced significant changes in efficiency scores for some companies, e.g. changes of up to 45 percentage points for Wessex Water in the WRP models compared to the PR19 LAD measure, which is unrealistically high. As population density is quite stable and predictable across years, it is not intuitive to observe such large changes produced by new alternative density drivers.

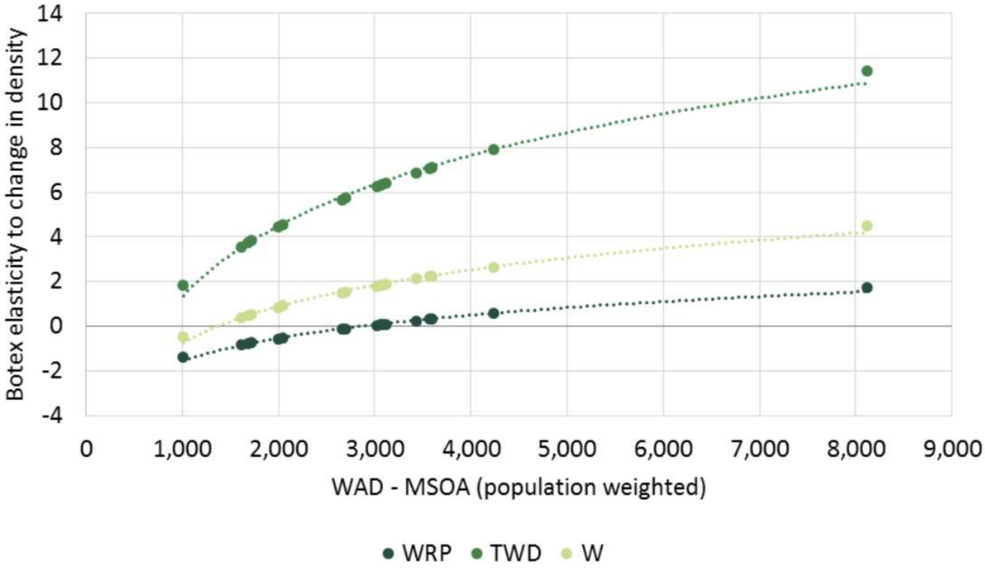
In addition, United Utilities (UU) has provided evidence that among the two WAD measures, WAD LAD from MSOA better aligns with the engineering rationale of a non-linear relationship between water base costs and density. This can be seen in the Figure 1.1 below, constructed by UU, which shows a U-shape relationship between WAD LAD from MSOA and each different elements of the water value chain, namely water resources plus, treated water distribution or wholesale water base costs. Conversely in Figure 1.2, we can see that there is no company with a negative elasticity between TWD costs and WAD from MSOA which suggests the inability of WAD from MSOA to capture the well-established U-shape relationship between density and TWD costs.

Figure 1.1 Botex elasticity to a change in WAD LAD from MSOA



Source: UUW (2023), 'UUW response - Consultation on econometric base cost models for PR24', page 12.

Figure 1.2 Botex elasticity to a change in WAD from MSOA



Source: UUW (2023), 'UUW response - Consultation on econometric base cost models for PR24', page 12.

While we acknowledge the concerns raised by a few companies on the aggregation of density at the LAD level which may change over time, we have also some concerns on the aggregation at the MSOA level because of the unexplained large changes in efficiency scores and its limited ability to reflect the U-shape relationship between density and costs.

However, we note that the majority of the industry supports WAD from MSOA and properties per length of mains length at the expense of WAD LAD from MSOA, since only five companies tend to prefer to rely on WAD LAD from MSOA. WAD from MSOA appears to be favoured only due to its greater level of granularity and its independence from LAD boundaries.

Unlike economies of scale at sewage treatment works or water network topography, where we benefit from a single cost driver that is clearly more appropriate and more statistically significant than any of the other alternatives (respectively WATS and APH), here it may be considered by Ofwat that there remains some uncertainty on the appropriate measure. As such, we consider that there may be merit in using different density measures.

If Ofwat considers that the PR19 approach regarding density can be further improved with the use of alternative density drivers, (i.e. the standalone use of WAD LAD from MSOA does not fully capture the impact between density and costs in the water modelling), then our view is that the second best approach is to complement WAD LAD from MSOA with WAD from MSOA. However, it is critical to keep WAD LAD from MSOA in the modelling suite - whether used alone, as in PR19, or triangulated with its PR24 alternative, WAD from MSOA.

In any case, we do not see why it would be appropriate to reintroduce properties per length of mains in the modelling when we now benefit from an exogenous, independent and reliable density estimate provided by the ONS. As mentioned in our May base cost consultation response, it was a good temporary solution at PR14 when we did not benefit from other robust alternatives, but this is not the case anymore, so this option is now outdated.

Indeed, properties per length of mains does not capture intra-zonal variations and the presence of sparse and dense sub-areas within a company's supply area. These types of arguments have been well made by SRN and TMS. Similarly, we would not expect Ofwat to continue to rely on properties per sewer length in the wastewater modelling for the same reasons stated above.

Finally, Ofwat should aim to achieve consistency across the three wholesale service areas and therefore we would not expect a density variable to be used in one service area but not in another, e.g. WAD LAD from MSOA in the water modelling and WAD from MSOA in the bioresources modelling.