PR24

NORTHUMBRIAN WATER living water | ESSEX&SUFFOLK WATER living water

RESPONSE TO CMA'S BASE COSTS WORKING PAPER

NON-CONFIDENTIAL

12 JANUARY 2026

STRUCTURE OF THIS RESPONSE

1 SUMMARY FOR THE CMA PANEL

1. We are responding to the CMA's Working Paper on base cost modelling (**Working Paper** or **WP**), with concerns, that if left unaddressed, will call into question the entire basis of the redetermination process.

2. The issuing of the Working Paper rightly recognises that 'a broad set of concerns were raised by stakeholders'[1] in relation to the experimental modelling proposed in the CMA's Provisional Determinations (**PD**). We welcome the CMA's willingness to engage further in this area.

3. However, the analysis and proposals detailed in the Working Paper are effectively limited to attempts to 'fix' the existing flawed approach set out in the PD and then only by examining a small subset of the (less material) problems we raised in response to the PD. It does not engage with the plausibility of the modelled cost allowances it presents. Nor does it seek to examine alternatives, which is fundamentally what is now required. We set out our concerns in this response and provide practical suggestions as to how the CMA panel can address these flaws in the limited time remaining in this process in order to reach a final determination that is consistent with the CMA's statutory duties. Fundamentally we consider that CMA must:

   - take a step back from the modelled results and consider what is the appropriate base cost allowance in the water sector context to avoid exacerbating the known base funding problems in the sector;
   - revert to the FD24 models or triangulate with other model results to improve stability and overfitting, better address the impact of energy prices in its cost allowances and reduce the over-reliance on a single model; and
   - move away from the upper quartile challenge, adopting a more appropriate 'median' challenge.

4. We believe that all these steps are necessary given the significant problems with the Working Paper's approach, its models and its proposed base cost allowances. It is essential that the CMA takes our suggested approach in order to reach a final determination that is consistent with the CMA's statutory duties and the principles of best regulatory practice. That would also reinforce the CMA's reputation as an expert body and reduce the risk of further challenge to the final determination.

1.1 BASE COST ALLOWANCES ARE FUNDAMENTAL TO THE DETERMINATION AND THE CMA SHOULD NOT MAKE THINGS WORSE FOR CUSTOMERS

5. Base costs represent the largest single element of the price determination and 72% of the bills

[1] CMA, Base Costs Modelling - Working Paper, 18 December 2025 (Working Paper), para. 2.6.

our customers pay. They provide the funding necessary to deliver the essential day-to-day services customers receive, ensure that we can maintain our assets in decent working order and that we are able to meet our core legal obligations.

6. We recognise that benchmarking of these costs can have valuable incentive properties as part of the regulatory framework. However, if these allowances are set too low, the consequences are likely to be a further deterioration in the health and condition of our assets and greater risk that we cannot deliver these essential services to our customers or meet our core legal obligations.[2] We are deeply concerned that the approach proposed by the Working Paper will inevitably have this effect:

- comparing the proposed AMP8 base allowances produced by the Working Paper models against actual base expenditure in AMP7, the Working Paper implies an immediate base cost reduction of c.9% (over £3.5bn) across the sector compared to what companies are spending now. This equates to c.£3bn of further efficiencies for the sector from the base costs allowed under FD24. This represents by some distance the largest base cost efficiency catch-up challenge that any reasonably comparable regulated sector has been asked to deliver in recent times;[3]
- during the most recent regulatory period every Water and Sewerage Company (**WaSC**) in the sector has overspent materially against their PR19 cost allowances. The aggregate overspend is equivalent to 14% of base cost allowances or around £5.4bn.[4] This experience has undoubtedly influenced the response to FD24, leading to the greatest number of appeals the sector has seen, all of which share the common thread of challenging the insufficient level of base cost allowances;
- external expert reviews of the sector, including by the Independent Water Commission (**IWC**), have concluded (or at least raised material concerns) that base cost allowances are (and have been) insufficient for capital maintenance.[5] The current replacement rates based on the most recent reported industry data are 770 years for water mains and 1,792 years for sewers.[6] These are clearly not sensible levels of investment;
- we have already shown the CMA that, despite being an efficient company with strong service performance, [redacted]

[2] See Letter from Wessex Water to CMA dated 18 December 2025 and Letter from Anglian Water to CMA dated 23 December 2025.
[3] See Figure 4 in Section 3.
[4] In 22/23 prices. The total overspend is c.£9.9bn on totex in current prices
[5] See NWL PD Response, footnotes 3 and 4 for references to commentary on asset health and investment from the IWC and the NAO.
[6] Sources: Companies Annual Performance Report Tables 6C and 7C for 2024/25. The total water mains replacement rate for 24/25 was 0.13% or 770 years. The total sewerage replacement rate was for 24/25 was 0.06% or 1,792 years. Sewerage includes Length of formerly private sewers and lateral drains (s105A sewers)

[redacted]; and

- our analysis of the PD found that the notional firm would barely be financeable to debt and equity. If the cost allowance in the CMA's final determination is as per the Working Paper, even a modest overspend would generate ratios inconsistent with the BBB+/Baa1 target for debt financeability. As such the notional firm would not be financeable under the Working Paper approach and a final determination that incorporates that approach would not satisfy the CMA's statutory duties.

7. There are scenarios where it is appropriate for a regulator to intervene in modelled cost allowances to ensure that companies are suitably incentivised to deliver significant efficiencies. But the approach and allowances envisaged by the Working Paper is wholly inappropriate in the current context of the water sector and by reference to our particular circumstances.

8. There is nothing sacred about the 'upper quartile' challenge which, in particular, is driving this outcome. It is best practise for regulators to cross-check any efficiency challenge with wider evidence and information. Base cost allowances would have been approximately £1.8bn higher over AMP7 across the sector if a median company benchmark had been used for PR19 when moving from econometric modelling results to setting allowances. This is highly relevant given the subsequent industry-wide overspends on AMP7 modelled base costs of £5.4bn. While there are interactions between allowances and expenditure levels it seems that regulatory judgements made by Ofwat and the CMA on purported catch-up efficiency explain about one third of the AMP7 over-spend. **We would urge the CMA panel to ensure that its redetermination does not repeat the mistakes of PR19 and move to a 'median' efficiency benchmark**.

9. On the balance of probabilities, if the CMA panel utilises the Working Paper approach in its final determination this would lead to gross overspending against allowances at a sector level and consequential impact on service performance for customers and the environment. It would further contribute to the long-term deterioration of the health of the asset base and the financeability of the sector. If we see another AMP of large base cost overspends across the sector this will reduce the credibility of the sector as an investment and drive up the cost of capital to the detriment of customers. In that scenario, the damage to the investability of the sector could easily become irreparable given the precarious state it is already in.

10. We implore the panel members to stand back from the models presented in the Working Paper and to consider the resulting proposed allowances in the real-world context of what is plausible, optimal and achievable.

1.2 THE WORKING PAPER DOES NOT 'FIX' THE CONCERNS RAISED WITH THE PD OR EXAMINE SENSIBLE ALTERNATIVES

11. We explained at length in our response to the PD why we did not consider LASSO to be appropriate for the PR24 redeterminations.[7] Our concerns with the use of LASSO were echoed in the responses from other stakeholders from all sides of this debate, including companies and Ofwat. This is not to say that techniques involving LASSO or cross-validation may not play some useful role in future regulatory reviews, but much more work is needed before they can be used reliability for setting critical allowances in the billions of pounds.

12. The Working Paper proposes some changes to its LASSO implementation and, in turn, to the resulting models. But these changes represent tweaks and adjustments that at best respond to more minor concerns while leaving the fundamental problems unaddressed. For instance, the Working Paper makes changes relating to "replicability" that tackle concerns that the modelling results are dependent on matters such as the ordering of candidate variables, but it does not make substantive changes to address concerns about the LASSO process leading to **over-fitting** and **instability of modelling results**. And while the Working Paper presents additional material on the robustness and sensitivity of the modelling results, **it crucially neglects to compare the stability of its models with the FD24 model suite** and shows no awareness of the ways in which its models perform more poorly than the FD24 models.

13. **A further critical error in the Working Paper's approach to model selection is seeking to select a single model** for wholesale water and a single model for wastewater network plus. The Working Paper does not seek to triangulate across a large and diverse set of models (as Ofwat does), to mitigate the risks that any single model will be imperfect and to mitigate sensitivity of allowances to subjective choices. There are inherent difficulties in seeking to use econometric models to capture the influence of external factors on water companies' base costs, especially within a relatively small data sample that exhibits limited variation over time in most of the explanatory variables and where explanatory variables are correlated. In this context, we consider it essential that allowances are set using triangulation across multiple models, that differ in significant aspects of their design. This is the established approach of Ofwat and the CMA.

14. **The inclusion of energy input prices in the modelling is an error.** While the CMA has sought to introduce energy prices into the modelling, the coefficients are both unstable across the Working Paper's models' variants and excessive. There is clear evidence from the data that the variable is picking up other factors that have driven up costs in recent years and are distorting the estimation of efficient costs for AMP8. This is consistent with energy costs only driving a small proportion of the AMP7 overspend and economic intuition about the link between the size of the

---

[7] NWL PD Response, Section 2.

expected coefficient and energy cost shares. Much more caution is needed about the treatment of energy prices, and it seems unsound to rely on this single approach.

15. Further consideration should be given to the inclusion of Average Pumping Head (APH) variable, the CMA must certainly explain why it considers the back-casted data is of sufficient quality. This is a material issue. We set out in Section 5.3 our substantial concerns with the inclusion of this variable just as others have done. The Working Paper provides a single sentence to note that the CMA does not see sufficient evidence to change its view about APH. The key issue is that the historical data cannot be back-cast accurately - this was a key finding of the Turner & Townsend review for Ofwat and we set out some of the many opaque adjustments that have been made to that historical data. If the historical data being used is not accurate then we fail to see how more accurate reporting now overcomes the problem that the CMA and Ofwat identified at PR19 where APH was excluded on data quality grounds.

1.3 PRACTICAL OPTIONS FOR THE CMA TO IMPROVE ITS MODEL SELECTION

16. We set out various practical proposals for the CMA panel to consider as it reflects on the responses to this Working Paper and moves towards its final determination. We consider that CMA must:

- revert to the FD24 models, which we still consider to be the best approach to base costs assessment available for use in this redetermination, or at least triangulate with other model results; and
- move away from the upper quartile challenge.

17. We believe that both steps are necessary.

1.3.1 Revert to FD24 models or at least triangulate results with other models focussing on those with greater stability

18. The Working Paper is highly problematic. It fails to address the fundamental problems with the approach taken in the PD. We see no valid way for the CMA panel to use the Working Paper models as the basis for setting allowances for AMP8.

19. Given the late stage in this process, we see a strong case for the CMA panel to revert to the FD24 models for its final determination. We show, for example, that the FD24 model suite is superior on the overall balance of evidence, taking account of key aspects of model performance. The FD24 models strike a far better balance between goodness of fit and the stability of results than the Working Paper's proposed models. Stability is critical as it shows that the results are not dependent on a single company and that the models represent genuine relationships between costs and drivers that apply across the whole sector. Overlooking this would be a significant mistake as the relationships are much more likely to be spurious and not reflective of genuine

differences in efficiency.

20. A decision to revert to the FD24 model suite for the CMA's final determinations would be informed by the CMA's exploration of LASSO-based models as a cross-check on the approach and the modelled allowances. This would be consistent with the CMA's objective to "explore a data-driven approach to variable selection using econometric tools such as LASSO" as a means to engage with specific claims put forward by some of the Disputing Companies. The CMA was clear that the use of such a data-driven approach to revise base cost allowances would depend on the "the robustness of the results and the extent to which they differ from Ofwat's".[8] As such, using this analysis to inform a decision to retain the FD24 models is, and always has been, a legitimate outcome of this exercise.

21. Alternatively, if the CMA decides to retain a role for the LASSO-derived models from its Working Paper, it will at the very least need to find a way to triangulate between these models and other, more robust models.

22. We do not see any sound basis for triangulation between the top-down and bottom-up LASSO models the Working Paper presents - this would simply perpetuate the weaknesses we have identified.

23. In simplest terms, the CMA might triangulate between the FD24 model suite and the Working Paper models. However, giving 50% weight to the Working Paper models would be excessive given the concerns about their provenance and quality.

24. Given this, we have carried out additional data-driven model optioneering that explores a much wider set of candidate models than the Working Paper's LASSO algorithm covers. From this exercise, we have identified a set of "frontier" models that offer the best goodness of fit for a given level of stability or the best stability for a given level of goodness of fit. These include models with stability closer to Ofwat's and models with goodness of fit closer to the Working Paper's models. A targeted selection of these models might provide a better basis for triangulation than the LASSO-derived models.

25. We have identified in the main body of our response how the CMA could use some or all of these frontier models as the basis for triangulation alongside the FD24 models. Figure 1 summarises our comparative assessment of how different options for model selection and triangulation perform in terms of addressing the three fundamental concerns we have identified with the Working Paper approach.

[8] CMA Consultation on Approach and Prioritisation, 28 May 2025, para. 43.

FIGURE 1: ASSESSMENT OF MODELLING APPROACHES AVAILABLE TO THE CMA

Model selection and triangulation option	Addressing concern about reliance on a single model	Addressing concern about over-fitting / instability of results	Addressing concern about CMA energy price modelling
Revert to FD24 model suite in light of analysis of wide range of feasible models	Broad triangulation	FD24 models strike a good balance of fit and stability	Energy prices adjustments applied outside the econometric models
Triangulation across FD24 models, LASSO models and "frontier" models	Broad triangulation (though need to avoid excessive weight to LASSO models)	Better balance of fit and stability than WP, but not as good as FD24	Triangulation would need to give sufficient weight to models that do not include the energy price variable
Triangulation across the set of "frontier" models from analysis of wide range of feasible models	Better than WP, but not as much diversity in models as under FD24	Better balance of fit and stability than WP, but not as good as FD24	Triangulation would need to give sufficient weight to models that do not include the energy price variable
LASSO model subject to removal of explanatory variables that are not statistically significant	Not addressed	Potentially, but may be at the expense of goodness of fit versus FD24 models	Addressed if energy price variables are dropped so that energy adjustments made outside the models
LASSO model with energy variable dropped and post-modelling energy adjustment	Not addressed	Not addressed	Energy prices adjustments applied outside the econometric models
Triangulation between LASSO-derived top-down and bottom-up models	Too few models & not enough diversity across them	Not addressed	Not addressed
Retain WP models (or similar LASSO-derived modes)	Not addressed	Not addressed	Not addressed

Source: NWL Analysis of PD and Working Paper

1.3.2 We see a strong case for a move to the median benchmark for base costs in AMP8

26. Using a median benchmark would be consistent with the approach generally taken by Ofwat and the CMA at PR24 to: (a) enhancement cost assessment; and (b) setting PCLs, where the median levels of performance over recent years are used for the PCL baselines for AMP8. For Ofwat, this reflects an informed decision to move away from the upper quartile benchmarks used at PR19 and in its DD24.

27. If the CMA does not consider that it should move away from the upper quartile benchmark then, at the very least, there is a very clear case for including a glide path as has been adopted in other regulatory contexts where efficiency challenges were clearly too severe.

28. However, given the experience of large overspends in AMP7, the expectations for substantial performance improvements over AMP8 simply to meet PCL baselines in the period to 2030, the serious concerns about asset health investment recognised by the IWC, and the need for a coherent and consistent approach across different parts of the price control determination, it is difficult to see how a rational regulator that has properly considered these issues would choose

anything materially more demanding than the median benchmark.

1.4 CMA MUST FOCUS ON THE EVIDENCE AND THE OVERRIDING OBJECTIVE NOT TRYING TO DISSUADE COMPANIES FROM FUTURE APPEALS

29. Finally, reflecting on the CMA’s role in conducting water company price control redeterminations, we recognise that the CMA has been vocal in expressing its view that these should be replaced with energy-style appeals in the future and that it finds them difficult. However, for the purpose of the PR24 redeterminations the legal framework remains clear – the CMA must, in stepping into the shoes of Ofwat, make a redetermination for each individual Disputing Company that best reflects the evidence and best meets the overarching objectives as set out in the s2 WIA 91 statutory duties, including inter alia the protection of current and future customers, the financeability of efficient companies and the resilience of the essential services.

30. In doing so the CMA should not be influenced by irrelevant factors, including for example the potential for its decision to encourage (or discourage) future challenges (noting the live debate about the future legislative framework for water redeterminations and that the position on a redetermination of PR24 for Thames Water remains, exceptionally and unusually, a live possibility even at this stage in AMP8) or the potential public reaction to its decision (reflecting the current negative sentiment expressed around the sector).

31. As has been made clear in our response to the PD and the Working Paper, we have found the CMA’s approach to base cost modelling perplexing. The wholesale rejection of Ofwat’s FD24 models in order to deal with a small number of targeted claims is a disproportionate response with material consequences that have not been properly considered. Furthermore, we do not understand why the CMA would be prepared to place such reliance on the novel LASSO-based approach when it comes to setting base cost allowances.

32. There are clear deficiencies in: the approach proposed in the Working Paper (e.g. the use of LASSO); the modelled cost allowances arrived at through use of the resultant models, which are implausible; and the process of engagement and consultation with the Main Parties and other stakeholders, which has been insufficient in duration and scope. The CMA panel must reflect carefully on the challenge and evidence presented in these responses, and the responses to the PD, and adapt its approach for the final determination accordingly. In that context, a decision to adopt the Working Paper approach in the CMA’s final determination would be unreasonable, irrational and open to challenge. We reserve the right to consider all options open to us if the CMA chooses to take that approach.

2 THE CMA'S PROPOSED ALLOWANCES FOR MODELLED BASE COSTS

33. In reaching its final determination, the CMA must have due regard to the context for the determination of base cost allowances. This needs to be understood as a judgement in light of econometric modelling and other evidence, including the real world context. This is not reflected in the Working Paper. In particular:

- we remain concerned that the approach put forward in the Working Paper has not been subject to an appropriate level of robust consultation and constructive development (see Section 2.2); and
- the CMA needs to be cognisant of the real-world implications of the material reduction in base cost allowances suggested by its proposed approach (see Section 2.2).

2.1 THE NATURE OF THE CMA'S APPROACH AND THE PROCESS IT HAS FOLLOWED

34. Whilst the CMA's decision to launch this supplementary consultation is positive, it is insufficient to address the material concerns we outlined in our response to the PD from both a process perspective and with regard to ensuring that the CMA's determination meets its statutory duties.[9] For instance:

- the changes proposed by the Working Paper which increase the scale of the reduction in AMP8 base cost allowances are even more disproportionate with respect to the claims raised by the Disputing Companies than the PDs were;
- our concerns about procedural unfairness remain – the FD24 cost models would not have been looked at by the CMA if this were a standalone redetermination for NWL. By opting to take this approach the CMA has gone beyond what is necessary to provide a targeted response to the specific points of dispute put forward by the Disputing Companies whilst simultaneously failing to properly test if the modelled results are credible across the sector;
- the limited tweaks made to the CMA's approach in the Working Paper have not adequately addressed our concerns about transparency and consistency;
- we continue to have material concerns about the quality of the modelled output and the accountability of this process, even with this additional phase of consultation, given its limited scope and lack of engagement with many of the points of challenge; and
- the CMA's approach remains inconsistent with regulatory precedent and the direction of travel from the IWC.

35. We have been consistent that a change of the magnitude posed by the replacement of Ofwat's base cost models must be subject to a suitable level of analysis, challenge, engagement and

[9] See, for example, NWL PD Response, paras. 13-15 and Section 2.1.

refinement. That has not been the case in this process. As such, it is not surprising that the CMA's use of LASSO and the chosen models in the PD and the Working Paper remain subject to such heavy criticism. In these circumstances it is neither reasonable nor rational for the CMA to pursue the Working Paper approach.

36. As we detail further in Section 2.2 below the Working Paper fails to engage with the plausibility of its modelled allowance in a real world context, including whether it might be feasible or optimal to achieve the level of base cost reductions the Working Paper envisages. This is of particular concern in light of the CMA's statutory duty to secure that each Disputing Company is able to finance the proper carrying out its functions. In that context, we would expect the CMA to be able to explain why the modelled costs it is presenting are appropriate. Instead, the CMA has failed to provide a coherent explanation for the reduction in base cost allowances under its approach, relative to Ofwat's (see Appendix 1).

2.2 REAL-WORLD CONTEXT FOR THE ALLOWANCES THAT THE CMA NEEDS TO DETERMINE

37. The allowances determined by the CMA have real-world implications. Ex ante regulatory allowances feed into budgets and act to constrain the scope for over-spends. They will affect what investments companies carry out and the operational resource available for provision of services to customers and protection of the environment. The CMA should be fully cognisant that its allowances will affect what companies spend and the value delivered for customers and the environment. As such, getting the base cost allowances right is fundamentally important to ensuring that the CMA meets its statutory duties and NWL is placed in a position to deliver against its statutory and regulatory obligations and meet customer expectations. The approach proposed under the Working Paper does not achieve this.

38. This is best demonstrated by looking at the implications of the modelled cost allowances for actual company expenditure, seen in the context of historical overspends and widespread concerns about underfunding of capital maintenance. This supports a reasonable forecast of overspends against the CMA's proposed allowance which will have negative implications for the financeability of the notional company.

2.2.1 The historic context for this redetermination

39. As has been set out in multiple submissions to the CMA (and discussed during the hearings) during the most recent regulatory period, AMP7, every company in the sector has overspent against its PR19 cost allowance. The average base totex overspend is equivalent to 13.5% of cost allowances or around £5.4bn.[10] Our total overspend in AMP7 was £372m.[11]

---

[10] Ofwat, WCPR-data-report-2024-25 wholesale base modelled data £4.6bn updated from 17/18 prices to 22/23 prices for consistency and comparability. The average base totex overspend is 11% if the impact of energy costs is removed.

[11] Ofwat, WCPR-data-report-2024-25. All Wholesale totex plus Retail overspends, £315m (17/18 prices) £372m (22/23 prices).

40. Whilst it has been suggested that this overspend could largely be attributable to an increase in energy prices, Ofwat has estimated that only 2% of this overspend was due to energy price increases, with the majority attributable to other factors, such as the costs of meeting a very stretching outcomes package.[12] This is consistent with the predictions made for AMP7 by the parties in the PR19 redetermination.[13]

41. Reflecting on these experiences, all the Disputing Companies raised concerns about the sufficiency of Ofwat's base cost allowances in FD24, making the case that Ofwat's assessment of the efficient level of base costs in AMP8 is not adequate to ensure an appropriate and sustainable level of funding.[14] For us in particular, this reflected the urgent need for increased spending on capital maintenance above the FD24 base allowances – a need recognised by Ofwat and the IWC.[15]

2.2.2 Magnitude of the base expenditure reductions under the CMA's Working Paper approach

42. In the Working Paper the CMA sets out details of the scale of its base cost allowance reduction against Ofwat's FD24 allowance.[16] These tables confirm that the Working Paper results in a reduction of the base cost allowance for all the Disputing Companies compared to the FD24 allowance. In our case, that reduction is particularly material (-8.9%). Across the sector the CMA finds that allowances should reduce by 7% compared to FD24 (over £2bn). Given that companies were already concerned that FD24 provided insufficient levels of base funding, the additional reduction makes that position worse.

43. However, that comparison simply indicates the relative position on costs between the CMA and Ofwat. To understand the real world implications it is informative to consider those allowances in the context of actual base expenditure that companies have incurred in AMP 7. That comparison is presented in Figure 2.

44. As Figure 2 shows, the Working Paper approach implies that in AMP8, with effect from year 1 of its redetermination, the sector can achieve a cost reduction of c.9% - equivalent to over £3.5bn. All the WaSCs would be expected to make material cost reductions, whereas it is only the Water Only Companies (**WoCs**) that see their expenditure increase. At the same time the companies are expected to make substantial improvements in service levels and environmental performance to avoid ODI penalties in AMP8.

---

[12] NWL PD Response, para. 131.
[13] See for example NWL's Statement of Case for the PR19 redetermination, Section 5; AWS SoC para. 4.
[14] See NWL SoC para. 15; ANS SoC para. 6; SEW SoC para. 1.11; SRN SoC para. 7; and WSX SoC para. 2.6.
[15] See, for instance, NWL Response to PD, Section 2.1.4 and IWC Final Report, Chapter 7. In the context of Ofwat, this is reflected in the adoption of the Asset Health Roadmap which accepts that it may be appropriate to allow additional expenditure on capital maintenance in AMP 8, over and above the FD24 allowance.
[16] Working Paper, Tables 3.4 and 4.3.

FIGURE 2: COMPARISON OF WORKING PAPER BASE COST ALLOWANCES FOR AMP8 WITH ACTUAL COMPANY BASE COST EXPENDITURE IN AMP7

Company	AMP7 actual base spend (£m, 2022-23 prices)	AMP8 WP base allowance (£m, post FS, 2022-23 prices)	Change AMP7 actual to AMP8 allowance (£m)	Change AMP7 actual to AMP8 allowance (%)
Affinity Water	1,171	1,226	55	4.7%
Anglian	**3,580**	**3,504**	**-76**	**-2.1%**
Bristol Water	403	412	9	2.3%
Northumbrian	**2,508**	**2,096**	**-411**	**-16.4%**
United Utilities	4,947	4,548	-398	-8.0%
Portsmouth Water	156	171	15	9.3%
SES Water	239	180	-59	-24.7%
South East	**782**	**802**	**20**	**2.5%**
Southern	**3,621**	**2,650**	**-971**	**-26.8%**
South Staffs Water	383	512	129	33.8%
Severn Trent Water (incl HDD)	5,312	5,245	-66	-1.2%
South West Water	1,897	1,494	-404	-21.3%
Thames Water	8,408	7,453	-956	-11.4%
Dŵr Cymru	2,533	2,438	-95	-3.8%
Wessex	**1,539**	**1,388**	**-151**	**-9.8%**
Yorkshire Water	3,548	3,393	-155	-4.4%
Total	**41,027**	**37,512**	**-3,514**	**-8.6%**
Disputing Companies	**12,029**	**10,440**	**-1,590**	**-13.2%**

Source: NWL analysis. AMP7 actual spend data is from Ofwat's updated datasets with actual 2024-25 data. AMP8 CMA updated allowances include frontier shift adjustment of 0.7%. NWL WPR Databook.

45. The position is particularly extreme for the Disputing Companies who collectively face a reduction of 13.2% (£1.6bn). We are facing an overnight reduction in our base expenditure of 16.4% (£411m). Southern Water is presented with a 26.8% reduction (£971m) and Wessex with a 9.8% reduction (£151m). If Thames Water were to trigger a redetermination now, it would be immediately facing an 11.4% reduction (£956m) in its base costs – a consideration which is likely to have a significant deterrent effect against exercising that option.

46. Such an outcome might be reasonable if the CMA had robust evidence to demonstrate that all WaSCs are inefficient and all WoCs are efficient or that certain companies are demonstrably more inefficient than others, or if it were confident that its modelling approach enables it to accurately identify efficient levels of costs. As we set out in this submission (and our PD Response) the modelling approach does not give rise to that level of confidence. Nor has the CMA presented any real-world evidence to support such assumptions about relative (in)efficiency.

47. Notably, the CMA has also failed to consider, in its PD or the Working Paper, whether this is an optimal outcome for the sector and customers, or if it is feasible for this level of cost reduction to be achieved in practice. Based on experience in previous AMPs, it is neither rational nor credible to make an assumption that the Disputing Companies can make this level of expenditure reductions without compromising customer service, environmental performance and without further deterioration in asset condition and system risk.

48. Indeed, even for companies that are broadly recognised as being efficient, the Working Paper approach assumes they would be capable of finding further cost savings. To take one example, Severn Trent is generally perceived as a particularly high-performing company in the sector (and the most efficient overall WaSC by Ofwat). While Severn Trent is anticipating some ODI rewards, it is not envisaging any under-spends against the Ofwat AMP8 allowances.[17] Ofwat's FD24 did not provide scope for Severn Trent to aim publicly for any degree of out-performance on base cost allowances. Yet the Working Paper modelling and allowances would indicate that Severn Trent's base expenditure should reduce by £66m. In contrast, the models suggest that the allowance for South Staffs Water should increase by 33.8% (£129m) but there is no consideration as to whether that would be optimal or in the best interests of its customers.

49. Rather than aligning allowances with efficient costs, it seems far more likely that setting allowances on the basis set out in the Working Paper will exacerbate problems in the sector. Further suppressing expenditure in AMP8 is likely to increase long term costs – which customers are directly exposed to via the established practice of 50% cost sharing – hold back performance improvements and build up risk in the system. Reducing base cost allowances so far below the levels from FD24 will conflict with the direction of travel from the IWC report which emphasis concerns about investment in capital maintenance and identifies a need for allowances to give more attention to operational reality and less to econometric modelling.[18]

50. We note that the impact on non-disputing companies should be a relevant consideration for the CMA when considering the plausibility of its approach, even though the direct impact is limited to the Disputing Companies. This is because in deciding how to address specific points of challenge raised by two of the five Disputing Companies the CMA has opted to completely replace Ofwat's base cost modelling approach (which is common across the sector) for all Disputing Companies (and potentially Thames Water). That suggests that the CMA's approach should be appropriate for any and all companies in the sector. As such, this situation is clearly distinguishable from the situation in the Bristol Water PR14 redetermination where the CMA focused on the impact of its determination and approach for the single company seeking a redetermination and noted that "we did not assess whether the approach that we used would be optimal or even feasible for Ofwat's periodic reviews of the price controls for all 18 water companies".[19] Whilst we agree that the CMA's focus should be on the individual companies whose price controls are being redetermined, we do not think it would be legitimate in the context of this PR24 redetermination to disregard whether its proposed models are 'optimal' or 'feasible' for all companies in the sector. Had the CMA's approach been adopted by Ofwat in FD24, we would expect to have seen all of the WaSCs

[17] Severn Trent Capital Markets Day, 5 March 2025
[18] IWC Final Report, Chapter 7.
[19] Bristol Water PR14 CMA Final Determination, para. 4.235(b). We note that the context of the CMA's approach to base costs in the PR14 redetermination was also different compared to PR24 – at PR14 the CMA looked at both econometric modelling and bottom-up engineering analysis and review of Bristol Water's business plan to sense check the modelling results.

seeking a redetermination – or at the very least giving it serious contemplation.

2.2.3 Practical implications for Northumbrian Water

51. In the PD Response we set out our concerns about the practical implications of the PD base cost allowances (corrected for the LASSO error) for us in AMP8. The magnitude of the challenge is increased under the modelled allowances presented in the Working Paper:

- on opex our analysis shows a gap of £38m a year to the Working Paper's base opex allowances, representing around 10% of controllable opex. This represents a 2% increase in the savings we need to find from FD24 (or £8m per annum). The £38m pa gap is equivalent to 90bps of RORE, pre cost sharing;[20]
- [redacted];[21] and
- [redacted].[22]

52. We do not repeat the substantive commentary in that response which applies equally to the results presented in this Working Paper but we do note that it is difficult to reconcile this position with our statement of case which compared cost efficiency and service performance information of the 10 WaSCs and showed us to be in the top quartile of performers. We were clear in the PD Response that managing a reduction in base cost allowances of this scale would have implications for our ability to manage the business and meet our regulatory and operational obligations. [redacted][23] This sub-optimal outcome receives no consideration in the Working Paper.

2.2.4 Implications for notional company financeability

53. As we noted in our PD Response persistent overspends of this magnitude over multiple AMPs (even if partly shared with customers) raises serious questions on financeability and our long-term cost of capital, and further casts doubt on Ofwat's claim that its base allowances are adequate to fund long-term capital maintenance expenditure requirements.

54. We demonstrate the impact on notional company financeability in Figure 3. If the CMA's final determination base cost allowance is based on the Working Paper approach and numbers and

[20] NWL PD Response, Section 2.1.4, para. 40 adjusted for Working Paper reductions.
[21] NWL PD Response, Section 2.1.4, para 42. adjusted for Working Paper reductions.
[22] NWL PD Response, Section 3.1.3, para. 170 and Figure 16, adjusted for Working Paper reductions.
[23] NWL PD Response, Section 3.1.3, para. 177.

we model a £200m (column WP-1) overspend to bring us up to Ofwat's base totex FD24 levels, this will result in financial ratios materially worse than those assessed by Ofwat in FD24. That overspend generates ratios inconsistent with the BBB+/Baa1 target. [redacted]

FIGURE 3: THE IMPACT ON FINANCIAL RATIOS OF A TOTEX OVERSPEND AGAINST WORKING PAPER ALLOWANCES

Ratios (5 yr average)	FD24	PD	WP-1	WP-2	BBB+/Baa1 threshold	BBB/Baa2 threshold
Notional Gearing	55.9%	55.6%	[redacted]	[redacted]	68%	75%
Adjusted cash interest cover (alt)	1.68	1.74	[redacted]	[redacted]	1.60x	1.40x
Funds from operations / debt (alt)	9.01%	9.43%	[redacted]	[redacted]	11.0%	8.0%

Source: Ofwat FD[24] and CMA PD[25] Financial Model Dashboard, NWL WPR Databook for CMA WP1 and CMA WP2, BBB+/Baa1 and BBB/Baa2 threshold data from NWL SOC Appendix 1.[26]

[24] FD24 NWL Financial Model, SOC322. Tab "Dashboard".
[25] CMA PD NWL Financial Model. Tab "Dashboard".
[26] NWL SoC Appendix 1, Section 7, Figure 24.

3 AN UPPER QUARTILE EFFICIENCY ADJUSTMENT IS NOT JUSTIFIED

55. In our SoC we articulated our view that it was not justified for Ofwat to persist with an "upper quartile efficiency adjustment" to reduce base costs allowances relative to the levels of costs predicted by its econometric models of historical expenditure. We noted that although this has been used for several price reviews, this is not, on balance, sensible given the limitations in the econometric modelling of base expenditure and risks a downward spiral in base allowances over successive controls.[27]

56. In our PD Response we articulated why an upper quartile adjustment should not be used where there are concerns about the quality and confidence of the underlying models.[28] Those concerns which applied to the PD models are equally applicable to the Working Paper models.[29] However, we consider this is an even more fundamental issue. This issue applies regardless of which set of models are used to set allowances; but it is an even greater concern in the context of the CMA's working paper modelling and the implications for AMP8 allowances.

57. We consider that upper quartile adjustments reflect outdated regulatory practice that is not grounded on reliable evidence on inefficiency. As the CMA acknowledged in its decision on the RIIO-2 price control appeals an upper quartile benchmark is not the default and may not be appropriate at every price review:

> Our decision should not be seen as indicating any preferred starting point for efficiency benchmarks. Regulators must always consider the case-specific circumstances and set the benchmark at a level appropriate for the case.[30]

58. Looking at the case-specific circumstances of this PR24 redetermination (such as the real world context detailed in Section 2.2) we see no rational basis for the CMA to make substantial deductions to modelled base costs on account of upper quartile "efficiency scores" from the econometric modelling. The CMA must take a step back and consider the balance of evidence, and balance of risk, in moving from the results of econometric modelling to setting allowances for AMP8.

59. In the following sections we set out the rational and evidence that supports a move away from the upper quartile adjustment from the CMA's final determination and the use of a median benchmark instead for base cost allowances over AMP8. In particular:

---

[27] NWL SoC, para 221 and Appendix 1, Section 5.3.
[28] NWL PD Response, Section 2.2.5 and Annex 1, Section 7.4. In particular we drew parallels with the CMA's approach in its PR14 redetermination where the CMA acknowledged the limitations in its modelling and concluded that an upper quartile efficiency benchmark would be "overly demanding": CMA PR14 Final Determination, para. 4.224.
[29] We also note that the Working Paper does not directly address the specific points of challenge we raised regarding the CMA's justification for an upper quartile benchmark in its PD.
[30] CMA, RIIO-2 Final Determination, Volume 3: Individual Grounds 28 October 2021 (SOC461) para. 12.142.

- it is misleading for the "efficiency scores" that are used to set the upper quartile adjustments being used as proxies for relative efficiency when, in fact, they are influenced by multiple other factors besides efficiency (see Section 3.1);
- upper quartile (or higher) adjustments have been made in previous AMPs and in other regulatory determinations, but the experience of material overspend across the sector in AMP7 demonstrates that the associated reductions to base cost allowances were overly demanding (see Section 3.2);
- in relation to ODI calibration, Ofwat learned from the experience of sector-wide under-performance and financial penalties during AMP7 and moved away from upper quartile performance benchmarks towards the use of median company benchmarks for PCLs. There is a need for the CMA to similarly learn from AMP7 and change the choice of benchmarks for base cost allowances (see Section 3.3); and
- Ofwat and the CMA have primarily used the median rather than upper quartile benchmarks in setting allowances for enhancement costs at PR24. It would be consistent to align the approach for base costs with the approach used for enhancements (see Section 3.4).

60. Noting the materiality of the cost reductions that would be required to meet the CMA's proposed base cost allowances (see Section 2.2) overspends are clearly inevitable if the upper quartile adjustment is made. Set in that context, and taking into account the other considerations outlined above, there is a clear regulatory rationale to take a different approach for this redetermination. We set out some suggested remedies for the CMA to consider in Section 3.5.

3.1 'EFFICIENCY SCORES' DO NOT SOLELY REFLECT EFFICIENCY LEVELS

61. The Working Paper refers to the "upper quartile (UQ) 'catch-up' efficiency challenge" and makes 28 references to "efficiency scores". This terminology follows that used by Ofwat but is misleading. The efficiency scores are not a reliable guide to differences in efficiency between companies. This is because they are influenced by other factors besides efficiency.

62. The "efficiency score" calculated by Ofwat and the CMA for a given company is obtained by dividing the company's actual base costs over the most recent five-year period by the base costs predicted by the econometric model(s) for that period.[31] These differences might reflect genuine differences in efficiency, but they are also highly likely to be affected other factors, such as:

- **limited ability to fully capture cost driver relationships**: there are inherent limitations in the ability of econometric benchmarking models, estimated using a small data sample, to: (a) capture all relevant external cost drivers; (b) for each cost driver, accurately capture the relationship between that driver and costs; and (c) take account of changes in cost driver

[31] More precisely, the comparison compares actual base costs (excluding certain costs that are treated as unmodelled costs and excluded from the base cost econometric models) against the predicted costs from the econometric models of base costs.

relationships over time;

- **costs differences related to performance differences:** there may be significant differences between companies' costs which reflect differences in levels of customer service and environmental performance which are not taken into account in the modelling. For instance, companies that have incurred less operational resource in the last five years would have lower efficiency scores but may not have performed as well as other companies in terms of customer service and environmental performance;
- **cost differences related to asset heath investment strategies**: there may be significant differences between companies' costs which reflect differing strategies towards asset health investment over the last five years. Companies with relatively low capital maintenance expenditure relative to other companies may be companies that have chosen to defer a greater amount of asset health investment; and
- **cost differences related to energy hedging position**: there may be differences between companies' costs during the five-years to 2023-24 which reflect good luck or bad luck in terms of companies' energy hedging and contractual positions, and the extent to which they were exposed to high energy price increases over this period.

63. These issues reflect the limitations in using econometric models to capture the influence of external factors on water companies' base costs, especially with a relatively small sample of companies. They also reflect the design of the econometric models, which do not seek to control for service quality and environmental performance, and do not take asset health into account. While we have particular concerns about the Working Paper's LASSO models (as set out in Sections 4 and 5), which give heighted risks of misinformation in relation to the efficiency differences between companies, these concerns apply more generally to the econometric modelling of base costs.

64. In the light of the issues above, to meet the costs implied by the upper quartile benchmark, companies would need to be even more efficient than an upper quartile company. As the CMA concluded in its decision for NIE in 2014:

> Weaknesses or limitations in the econometric models and any errors or inconsistencies in the data set we used will contribute to the variance in costs across the 15 companies in the sample. We would expect this to have an effect on the statistical properties of the cost benchmarks. We would expect this variance to introduce a bias that overstates the relative performance of companies ranked better than the median performance and understates the relative performance of companies ranked worse than the median. Where we see a company that has performed relatively well in the benchmarking analysis we would expect that, on the balance of probability, its performance or rank has been improved (to some degree) by modelling limitations and data issues.
>
> In the presence of modelling limitations and data error, we expect that our choice of the fifth company for the benchmark means that, on the balance of probability, NIE would need to be more efficient than the

fifth company if its costs are to match our estimated cost benchmark. An effect of modelling limitations and data issues is that the cost benchmark is more demanding than it might appear.[32]

65. On the cost/performance relationship, there has been a tendency by Ofwat, and in some cases the CMA, to reject concerns about these impacts in the absence of compelling evidence. It is wrong to assume that the gap between modelled costs and actual costs is due to efficiency unless proven otherwise.

66. Similarly, it is not appropriate to dismiss the significance of companies' decisions about capital maintenance for the efficiency scores. The cost/asset health relationship is not about capital maintenance "troughs" and cannot be assessed by simply looking for cyclical patterns in individual companies' expenditure. This might be better demonstrated if there were much better data on asset health across companies and over time. But given the data that is available there is no assurance that companies with lower efficiency scores are not companies that have been carrying out less long-term asset health investment and building up more risks in their systems.

67. As set out in further detail in the Economic Insight report[33] there is evidence of significant differences between companies in terms of energy hedging and power procurement contractual positions around the time of the energy crisis. This will lead to differences in efficiency scores for reasons relating in part to luck rather than efficiency.

3.2 EVIDENCE FROM AMP7 DEMONSTRATES THAT THE EFFICIENCY ADJUSTMENTS APPLIED TO BASE COSTS AT PR19 WERE TOO DEMANDING

68. The application of an upper quartile (or similar) efficiency adjustment for base costs is familiar from water company price control determinations. But it remains a regulatory judgment with substantial financial implications and a fresh assessment is needed at each review.

69. Figure 4 demonstrates that while regulators have used an upper quartile benchmark for base costs in previous price control decisions, there have been variations in the size of reduction applied.

FIGURE 4: REGULATORY PRECEDENT ON THE APPROACH TO CATCH UP EFFICIENCY ADJUSTMENTS

Precedent	Efficiency benchmark	Reduction applied to costs predicted by the econometric models
Ofwat PR14[34]	Upper quartile	Water: 6.5%; Wastewater: 10.4%
CMA PR14[35]	Average	Water: 0%
PR19 Ofwat[36]	4th & 3rd most efficient company	Water: 4.6%; Wastewater: 2%

[32] Competition Commission, NIE Final Determination (2014) para 8.135-8.1366.
[33] Economic Insight, The Treatment of Energy Input Price Inflation in Base Cost Econometric Models, 9 January 2026 (**EI Energy Report**) (**NWL-WPR-001**), Section 3. This report has been prepared on behalf of all Disputing Companies.
[34] Ofwat, PR14 FD policy chapter A3 – wholesale water and wastewater costs and revenues, p. 4; CMA PD Table 4.3, p. 56.
[35] CMA, Bristol Water Final Determination PR14, para 4.245, p.123.
[36] CMA PD Table 4.3, p. 56.

PR19 CMA[37]	Upper quartile	Water: 1.4%; Wastewater: 3.3%
RIIO-GD2[38]	Upper quartile to 85th percentile glide path	3.5% to 3.9%
PR24 Ofwat[39]	Upper quartile	Water: 1.3%; Wastewater: 0.6%
PR24 CMA PD[40]	Upper quartile	Water: 5.6%; Wastewater: 4%
PR24 CMA WP[41]	Upper quartile	Water: **7%**; Wastewater: **6%**

Source: As per footnote references

70. As reflected in Figure 4 at PR14 the CMA used an average efficiency benchmark for Bristol Water. The CMA also explained at PR14 that it had used its more bottom-up review of Bristol Water's business plan as a "cross-check" on the econometric benchmarking analysis, citing concerns that the "*econometric benchmarking analysis might not be able to take sufficient account of Bristol Water's needs and circumstances*".[42]

71. We can see that the adjustments proposed under the CMA's working paper are extremely high relative to historical precedent. Furthermore, they are untenable in the broader context, particularly given the AMP7 overspends.

72. Figure 4 shows the adjustments made to modelled costs in light of the calculation of efficiency scores. In practice, applying an adjustment to predicted costs based on efficiency scores calculated over the last five years (the approach of Ofwat and the CMA for PR24 and previous review) leads to an adjustment that reflects two different things: (a) differences between companies' costs over the last five years versus the full sample period; and (b) differences in costs between companies (over the last five years).

73. Adjusting for the first factor is an absolutely necessary adjustment to ensure that allowances for the forthcoming price control period are set using more recent expenditure data that better reflects the costs over that period.[43] This is especially so given that the econometric models do not take account of differences in costs over time that are not explained by changes in the various cost driver variables. While there are some fluctuations in costs between years related to the price control cycles, companies' base costs have generally increased over time to meet increased performance and compliance expectations and as a result of operating a larger set of assets.

[37] CMA, PR19 Final Determination Financial Model WW1, tab "Efficiency"; CMA, PR19 Final Determination Financial Model WWW2, tab "Efficiency".
[38] Ofgem, Ofgem RIIO-3 Final Determination Technical Annexes, file "CostAssessment_File", tab "RIIO2 Cal_Efficiency", row 103-104.
[39] CMA PD, Table 4.3, page 56.
[40] CMA PD, Table 4.3, page 56.
[41] Working Paper, para 3.20 & para 4.18.
[42] CMA, Bristol Water Final Determination PR14, para. 5.5.
[43] This is needed at the minimum. But it is not enough. As set out in our SoC (see Section 4.2 and Appendix 1 Section 4) and PD Response (see Section 3.2) Ofwat and the CMA's overall cost assessment relies too heavily on historical evidence and lacks sufficient adjustments for factors causing costs to increase over the forthcoming control period beyond the level of costs in the most recent part of the historical data.

74. The second aspect of the adjustment – the adjustment for differences in efficiency scores between companies – is far more problematic. However, the extent of this problem is somewhat obscured by the efficiency scores conflating the intertemporal effects above with cross-company effects.

75. In Figure 5 below, we disentangle these issues by showing how much more demanding the catch-up efficiency adjustments applied by both Ofwat and the CMA at PR19 were than if these had been set on the basis of a median company benchmark (e.g. as was used for enhancement cost benchmarking at PR19).

FIGURE 5: IMPACT OF USING EFFICIENCY BENCHMARKS MORE DEMANDING THAN THE MEDIAN AT PR19 FOR WATER AND WASTEWATER

	Ofwat PR19		CMA PR19	
	Water	**Wastewater**	**Water**	**Wastewater**
Efficiency score under benchmark used for determinations	0.95	0.98	0.99	0.98
Median efficiency score	1.01	1.02	1.03	1.02
Additional cost reduction from using PR19 UQ(+) benchmarks rather than the median cost benchmark	**5.2%**	**4.0%**	**4.7%**	**3.8%**

Source: PR19 benchmark and median efficiency scores from Ofwat FD and the CMA feeder model 2. NWL WPR Databook.

76. The cost reductions applied at PR19 as shown above, stemming from regulatory decisions to apply very challenging efficiency adjustments rather than using the median benchmark, are highly significant when it comes to understanding the industry's outturn expenditure over AMP7.

77. As set out in Section 2.2.1, during AMP7 companies systematically overspent against their modelled base cost allowances, by 13.5% on average. This is a very large over-spend and a serious concern for the regulatory regime.

78. Of this 13.5%, Ofwat has estimated that of the AMP7 overspend, 2% is due to the effects of unexpected energy price increases (in excess of CPIH).[44] Economic Insight have carried out further analysis and found that "only a minority (between around a quarter and under a third) of the AMP7 overspend can be explained by the energy cost shock".[45]

79. Neither Ofwat nor the CMA have provided good explanations for the large residual base cost overspend in AMP7. Indeed, neither Ofwat nor the CMA have shown sufficient interest in understanding what might have caused that overspend and whether there are implications for regulatory decision-making at PR24. This is an especially important omission for the CMA's redetermination now that we have outturn expenditure data across the whole of AMP8. Ofwat has pointed to the leakage performance challenge over AMP7 as another factor contributing to the overspend. [46] But this does not explain the magnitude of the overspend and, in any event, across

[44] Ofwat PR24 redeterminations Expenditure allowances – addressing asset health, April 2025, p. 17.
[45] EI Energy Report (NWL-WPR-001), p. 5.
[46] Ofwat PR24 redeterminations Expenditure allowances – addressing asset health, April 2025, p.18.

the package of financial ODIs set for AMP8 the PCL baselines require substantial performance improvements to be achieved from base cost allowances, so this is not an issue that has gone away.

80. Clearly one key driver of the industry-wide over-spend on base costs is the choice of efficiency benchmark. Had Ofwat and the CMA used the median value from the efficiency scores to adjust predicted costs, base cost allowances for companies would have been around 4%-5% higher over AMP7.

81. We do not see how the CMA can avoid the conclusion that the application of demanding reductions to base costs, of the order of 4-5% relative to the use of a median company benchmark as shown above, is a key factor that led to water companies over-spending allowances at PR19.

82. Across the industry, base cost allowances would have been approximately **£1.8bn**[47] higher over AMP7 if a median company benchmark had been used when moving from econometric modelling results to setting allowances. This is highly relevant given the subsequent industry-wide overspends on modelled base costs over AMP7 of **£5.4bn**.[48] While there are interactions between allowances and expenditure levels it seems that regulatory judgements made by Ofwat and the CMA on purported catch-up efficiency explain about one third of the AMP7 over-spend.

83. Leaving aside concerns about the emphasis placed on econometric models (as expressed in the IWC final report), and debates about the choice of what models to use, both Ofwat and the CMA made subjective judgements at PR19 to apply reductions to the costs "predicted" by those models calculating using upper quartile (or similar) benchmarks. With the benefit of hindsight, and in view of industry-wide expenditure and ODI performance, those reductions were misguided.

84. The Working Paper provided an opportunity for the CMA to give proper consideration to the choice of benchmark in light of the latest modelling results. Instead, the Working Paper's persistence with upper quartile adjustments, combined with the changes to its econometric models, mean that there is an even greater problem for base cost allowances in AMP8. In the table below, we show that under the Working Paper's proposed approach there would be significant reductions to predicted costs in setting allowances compared to using a median benchmark.

FIGURE 6: EXTRA COST REDUCTION FROM UQ VS MEDIAN UNDER WORKING PAPER MODELS

	Water	Wastewater
UQ efficiency score under WP models	0.93	0.94
Median efficiency score under WP models	1.01	0.97
Extra cost reduction from using UQ benchmarks rather than the median	**8.5%**	**2.9%**

Source: Working Paper; FD24, NWL Analysis

[47] NWL WPR Databook for Figure 5
[48] Ofwat, WCPR-data-report-2024-25 wholesale base modelled data £4.6bn updated from 17/18 prices to 22/23 prices for consistency and comparability.

85. The extra cost reduction from using the upper quartile rather than median is greater under the Working Paper modelling than under the FD24 model suite raising even greater concerns.

3.3 OFWAT AND THE CMA HAVE ADAPTED THE APPROACH TO PCLS IN LIGHT OF EVIDENCE ON PERFORMANCE ACROSS THE SECTOR

86. There is highly valuable insight and guidance for the CMA to gain from considering how the regulatory framework has developed for water company ODIs.

87. In setting PCLs for AMP8 (i.e. the performance baselines above which a reward is earned and/or below which a penalty applies), Ofwat responded to evidence on the widespread under-performance across the sector during AMP7 in relation to common ODIs.

88. At PR19, there was considerable emphasis on upper quartile performance when calibrating PCLs for common PCs.[49] For three PCs with common PCLs across the sector (water supply interruptions, internal sewer flooding and total pollution incidents) Ofwat set PCLs based on projections of upper quartile performance over AMP7 (with some moderation in the case of water supply interruptions).[50]

89. In FD24 Ofwat largely moved away from a position that PCLs should be calibrated by reference to upper quartile performance benchmarks. This was in light of evidence of extensive under-performance on ODIs.

90. In practice, for most PCs with common PCLs set using historical performance data, Ofwat chose at PR24 to use the median level of historical performance to set the 2024/25 PCL baselines, which in turn grounded the PCLs across AMP8 by reference to recent median performance. In explaining its change in approach Ofwat said:

> The outcomes package aims to reward good performance and penalise poor performance. We do so by incentivising companies to aim for their PCLs and go beyond them. If the PCLs are beyond what is achievable by the majority of companies, we are not giving the sector a fair chance to earn rewards from improved performance. This runs the risk that companies might decide it is not worthwhile to the make the effort to pursue their PCLs, which will not benefit customers or the environment. It would also result in the sector being placed at high likelihood of immediate underperformance from the start of the 2025-30 period. This would also not support delivery of the large investment programme required to deliver improvements for customers and the environment over the 2025-30 period and beyond.[51]

91. This was a late change which made a material difference to the overall outcomes package. The CMA has adopted this approach and taken it further. For example, at PR19 Ofwat and the CMA choose to use upper quartile performance benchmarks to set PCLs for water supply interruptions.

---

[49] Ofwat PR19 Final Determinations: delivering outcomes for customers policy appendix, December 2019.
[50] CMA PR19 Final Determination para 7.143.
[51] Ofwat FD24 PR24 final determinations: Delivering outcomes for customers and the environment, p. 21.

In its PD the CMA has rightly applied Ofwat's updated approach to PCL calibration more consistently, such that PCLs for water supply interruptions over AMP8 are grounded on **median levels of performance** over the four-year period to 2023/24.

3.4 ENHANCEMENT ALLOWANCES AT PR24 ARE PRIMARILY SET USING MEDIAN BENCHMARKS

92. Looking across the approach to enhancement cost assessment for the PR24 review, almost all areas of cost were set by Ofwat with reference to a median company efficiency score benchmark, not upper quartile (see Figure 7).

FIGURE 7 - OFWAT PR24 APPROACH TO CATCH-UP EFFICIENCY ADJUSTMENTS FOR ENHANCEMENTS

FD24	Approach to catch-up efficiency
Enhancement models:	
Storm overflows	Median (using blend of historical and forward looking)
Phosphorus	Median (using blend of historical and forward looking)
Growth	Median
IED	Upper quartile for secondary containment, median for tank covering
Sanitary parameters	Median
Interconnectors	Median (using blend of historical and forward looking)
Metering	Median
Lead replacement	Median
All unit cost models	Median

Source: Ofwat FD24 Enhancement Cost Modelling Appendix, p.4-9

93. Figure 7 shows that in FD24 Ofwat uses a median company challenge for *all* enhancement models except for one element of IED (secondary containment) where Ofwat had specific reasons for a stronger efficiency challenge and applied a 25:25 cost sharing rate as a result.[52] Where Ofwat did not use models, it instead used median unit costs to set catch-up efficiency targets.

94. For storm overflows and phosphorus Ofwat says that it did "not apply a more stringent efficiency challenge as we consider the use of historical models leads to a sufficiently stretching but achievable cost challenge".[53] This efficiency challenge was equivalent to median for storm overflows (as it directly uses the values derived from models). If Ofwat had applied an upper quartile efficiency challenge, this would have meant up to a further 7% reduction in totex allowances.

95. In the PD the CMA considered options of median and upper quartile benchmarks for p-removal costs as well as the option of no such adjustments (i.e. using predicted costs from the benchmarking directly). It chose the third of these options which was the least challenging in terms of allowances, but still involved lower allowances than companies had requested:

[52] Ofwat FD24 Enhancement Cost Modelling Appendix, p.123.
[53] Ofwat DD24 Enhancement Cost Modelling Appendix, p.3.

In line with the approach taken by Ofwat in its PR24 FD modelling of enhancement expenditure, we considered setting a median efficiency challenge for p-removal allowances. In our model, the median efficiency score before outlier adjustments are applied for future schemes funded in AMP8 is 0.97 and when applied sector wide would result in 5.5% reduction in totex compared to the amount requested. We also considered setting an upper quartile efficiency challenge. The upper quartile efficiency score is 0.95 in our model and would result in the sector receiving 7.7% less totex than requested.

Given that our model awards 2.6% less totex than the sector requested without further adjustments to modelled costs prediction, we have provisionally decided not to apply any efficiency challenges to p-removal enhancement allowances.[54]

3.5 SUGGESTIONS FOR THE CMA'S FINAL DETERMINATION

96. It is entirely right that Ofwat and the CMA have learned from the experience of AMP7 and made major changes to the way that ODIs and PCLs are calibrated to (in Ofwat's words) give "the sector a fair chance to earn rewards from improved performance" and avoid a "high likelihood of immediate underperformance from the start of the 2025-30 period".[55]

97. Is now essential that the CMA makes a corresponding change for base cost allowances, moving away from the upper quartile. For consistency across the calibration of PCLs and base cost allowances, this could involve the following:

- set the starting position for AMP8 allowances based on a median rather than upper quartile benchmark; and
- continue to apply a frontier-shift efficiency improvement factor over time as per the PD.

98. We consider that this move to the median benchmark should be made regardless of whether the Working Paper or FD24 models (or a combination thereof) are used. A regulatory response to the AMP7 over-spend is needed regardless of the models. But given the implications for allowances (and the issues discussed in Section 2) there is an even greater need for this change if the CMA uses the Working Paper models to set allowances for AMP8.

99. This would avoid material and inevitable (repeated) overspends versus allowances and consequential impacts on service performance and asset health for customers and the environment. It would also rightly recognise the inherent weaknesses with cost models and respond to the real-world evidence of overspending in the previous period, learning from the mistakes of PR19 as Ofwat and the CMA have done on enhancement cost assessment and ODIs.

100. This is not the only approach that might be considered in principle. For instance, between the upper quartile benchmark and median benchmark, there are a range of other benchmarks that

[54] CMA PD, para 5.86-5.87
[55] Ofwat FD24 PR24 final determinations: Delivering outcomes for customers and the environment, p. 19.

might be considered, such as:

- the median or average efficiency score across all companies excluding companies with efficiency scores that are at the lower quartile or worse (e.g. if there were concerns about influence of any highly inefficient companies on AMP8 allowances across the industry);
- a judgement-based adjustment so that only a proportion of the cost reduction that would apply from the upper quartile efficiency score is used to reduce allowances below modelled costs (with judgment reflecting the influence of other factors);
- a benchmark based on the 40th percentile or upper third positions for efficiency scores; or
- a glide path from median to upper quartile performance.

101. However, given the experience of large over-spends from AMP7, the expectations for substantial performance improvements over AMP8 simply to meet PCL baselines in the period to 2030, the serious concerns about asset health investment recognised by the IWC, and the need for a coherent and consistent approach across different parts of the price control determination, it is difficult to see how a rational regulator that has properly considered these issues would choose anything materially more demanding than the median benchmark.

4 THE WORKING PAPER MODEL SELECTION PROCESS

102. We have explained at length why we do not consider LASSO to be appropriate for the PR24 redeterminations.[56] Our concerns with the use of LASSO were echoed in the responses from other companies and Ofwat. Techniques involving LASSO or cross-validation may play some useful role in future regulatory reviews, but much more work is needed before they can be used reliably for setting allowances in the billions of pounds to support the provision of essential services. The PR24 redeterminations should not be the place for econometric experimentation.

103. The Working Paper has made some changes to the LASSO implementation and, in turn, to the resulting models. But these changes represent tweaks that at best respond to more minor concerns while leaving the fundamental problems unaddressed.

104. LASSO remains a novel and experimental model selection technique for water company base cost benchmarking purposes and nothing in the Working Paper provides any comfort that the scale of methodological departure from established regulatory practice is acceptable for the PR24 redeterminations. We focus on three aspects of the Working Paper model selection process that are deeply problematic:

- the Working Paper's LASSO process does not provide effective protection against over-fitting (see Section 4.1)
- failure to adequately compare the robustness of the proposed models against alternatives (see Section 4.2); and
- reliance on a single econometric model for each of water and wastewater (see Section 4.3).

105. To inform our response to the Working Paper, we engaged Melvyn Weeks, an Associate Professor of Economics at the University of Cambridge, to assess the updated approach.[57] His research interest spans both theoretical and applied microeconometrics including policy evaluation, model testing and evaluation; and computationally intensive methods including machine learning, simulation-based inference and the bootstrap. The LASSO technique falls within the machine learning area where he also teaches at post graduate level. We have referred to that assessment in this response and encourage the CMA to consider Professor Weeks' report.

106. Further to the three fundamental concerns above, and as noted by Professor Weeks,[58] the implementation of the LASSO approach means that highly-complex and opaque user choices about the choice of algorithm or approach to cross validation can affect the models selected by the LASSO process and the expenditure allowances for individual companies (as we have seen in the large differences between the CMA's PD and Working Paper modelling results). This

[56] NWL PD Response, Section 2.
[57] Professor Melvyn Weeks, CMA Consultation on Base Cost: A Response, 11 January 2025 (**Weeks Report**) (**NWL-WPR-002**).
[58] Weeks Report, NWL-WPR-002, p.ii.

reinforces the view that this is not a technique that can be deployed for PR24 in place of the FD24 modelling. The remainder of this section turns to the more fundamental concerns with the Working Paper's approach.

4.1 THE LASSO PROCESS DOES NOT HAVE EFFECTIVE PROTECTION AGAINST OVER-FITTING

107. There is a major risk with a data-driven process such as LASSO that the model selection process becomes one of selecting the model that, amongst those models identified and estimated, has the best in-sample goodness of fit but is actually capturing spurious correlations in the data. It is well-known that simply maximising goodness of fit does not produce an accurate and reliable model.

108. The Working Paper's LASSO process might provide some protection against that risk through the use of cross validation and the choice of the fraction of the L1-norm parameter which means that the selected model is not necessarily the model with the lowest in-sample RMSE. However, there is simply no guarantee that this will be effective in the context of the base cost modelling.

109. We found that the PD models were subject to over-fitting.[59] The Working Paper makes some technical changes to its LASSO implementation in response to concerns about the reproducibility of its model selection results.[60] While the reproducibility issue seems to have been tackled, the problem of over-fitting under the LASSO process remains.

110. One aspect of this concerns the choice on the value of the fraction of the L1-norm used. This parameter takes a value between 0 and 1, with lower values imposing a higher penalty on the inclusion of a greater number of cost drivers in a model. It plays a role akin to the "lambda" parameter used in the LASSO algorithm that the CMA drew on in its PD analysis. The Working Paper, for each of the models, uses a rule that the value of that fraction should be the value that minimises the mean square prediction error of its cross-validation. In doing so, the CMA has cast aside the concerns it raised at PD concerning the risk of over-fitting from selecting a value of the penalty parameter that minimised the square error of the cross-validation.[61] That concern led the CMA to put forward its preference for using "lambda.1se" in the PD. It does not seek to take an analogous approach now, though the concern on the risk of over-fitting remains.

111. A second aspect concerns the approach to cross-validation. The Working Paper's LASSO process involves cross-validation as part of the defence against over-fitting. In principle, and taken broadly, cross-validation might help reduce risks of over-fitting by creating differences between the dataset that the model is estimated on and the dataset used to assess the goodness of fit of model estimation results. However, for the Working Paper, the approach to cross-validation is Leave-

[59] NWL PD Response, para 102.
[60] NWL PD Response, Section 2.2.4.
[61] PD, para 4.55

One-Out Cross-Validation (**LOOCV**). In the CMA's words, this method: "leaves out one observation at a time, fitting the model to the remaining data and evaluating predictive performance".[62]

112. This is a relatively weak approach to cross validation, especially in the context of a panel of water companies over time. The panel of 17 water companies over 13 years provides 221 observations for wholesale water modelling. It is a weak approach to cross validation to estimate the model on all but one of these observations (i.e. to estimate on 220 rather than 221 observations). Professor Weeks' report sets out that:

> A more appropriate cross-validation procedure for panel data in regulatory contexts is Leave-One-Company-Out (LOCO). LOCO directly tests the regulatory question by withholding all observations for a given company and predicting that company's costs using only information from other companies.
>
> If cross-validation error under LOCO substantially exceeds that under observation-level LOOCV, this would indicate that the CMA's reported prediction accuracy overstates the model's ability to serve its regulatory purpose. The absence of this analysis means stakeholders cannot assess whether the methodology's apparent precision reflects genuine predictive capability or merely an artefact of the validation design.[63]

113. A stronger and more suitable approach to cross validation, to help protect against over-fitting, would involve the exercise of looking at dropping whole companies and/or years.

114. But responding to the issues above cannot salvage the Working Paper's general approach to model selection. For the specific dataset and set of candidate variables under consideration, the LASSO process is not geared towards estimating and testing a sufficient range of feasible models as to be able to select models that strike a good balance between goodness of fit and stability. LASSO is the wrong choice of approach for the purposes of PR24 base cost benchmarking, even if it might have value in very different situations. Professor Weeks gives the example of using LASSO for the very different exercise of "screening thousands of genetic markers to identify a few that predict disease outcomes".[64]

115. As discussed in Section 6 we find that there are other feasible models which can provide similar levels of goodness of fit (on an RMSE basis) with much better protection against over-fitting risks (as evidenced in greater stability of estimation results, fewer selected variables and better statistical significance). We also find that the FD24 model suite provides a better balance of goodness of fit and stability than the LASSO-derived models.

---

[62] Working Paper, para A.8
[63] Weeks Report, NWL-WPR-002, p.6.
[64] Weeks Report, NWL-WPR-002, p.i.

4.2 FAILURE TO COMPARE THE ROBUSTNESS OF PROPOSED MODELS AGAINST ALTERNATIVES

116. The fifth and final element of the Working Paper's response to the concerns raised with the PD is what the CMA describes as "Robustness / Sensitivity Checks", which are applied after the LASSO model selection process.[65] This comes across as a superficial tick-box exercise as there are a number of limitations to the approach presented in the Working Paper.

117. Most importantly, there is no proper evaluation of whether the Working Paper's models perform better or worse than the FD24 models in terms of model robustness (or against other feasible models beyond the model selected via the LASSO process). The Working Paper includes some analysis of the stability / sensitivity of its modelling results to changes in the data (using a form of bootstrap resampling). It has not compared the sensitivity of the LASSO-derived models to the FD24 model structure. In addition, the Working Paper's approach to the bootstrap resampling used by the CMA does not seem well-suited to the panel data structure. We discuss the assessment of model robustness further in Section 5.1.

118. While the CMA said that it has subjected its models to the suite of statistical tests routinely applied by Ofwat, it does not adapt the Working Paper models when they models fall short against these tests. This means that the Working Paper does not make any changes despite its model selection process leading to models with variables not statistically significant even at the 10% level, whereas Ofwat treated such t-tests as highly important. Nor has the CMA applied all of the model robustness tests applied by Ofwat, in particular omitting the tests on sensitivity to dropping companies and years.

119. Furthermore, the Working Paper only considers one model for water and one model for wastewater when carrying out these tests when given the pool of 8 candidate variables for water and wastewater there are a total of 255 possible models for each. The chosen LASSO algorithm only selects one preferred model according to the user choices in programming it to focus on minimising the RMSE. However, it does not explore other models that may perform better in other dimensions such on the statistical significance of the variables, the performance against the diagnostic tests, or the stability of the results.

120. The Working Paper's approach has potential benefits in the context of large datasets with a large number of candidate explanatory variables, especially where many of those candidate explanatory variables have no material influence on the subject of the analysis. That is not the case in the current context as a) we have a small panel dataset where most of the variation is between companies, and b) we are starting from a pre-selected pool of likely relevant variables.

121. Melvyn Weeks explained as follows in his updated assessment of the Working Paper approach:

[65] Working Paper, Figure 2.1

> LASSO is a variable selection method premised on the sparsity assumption: that the true data generating process involves only subset of relevant predictors, with some coefficients being exactly zero. […] This approach was developed for high-dimensional settings where the number of potential predictors p substantially exceeds the sample size n.[66]

122. This is far from the world of water cost modelling assessment. It is essential that the CMA considers more models to make an informed decision over which better meets to the relevant objectives of ensuring customers are protected and that companies can carry out their functions. The failure to properly consider alternatives is indefensible for the purposes of selecting models for base cost benchmarking in the PR24 redetermination.

123. Overall, the Working Paper's approach to checking model robustness is inadequate, especially given the need to test the outputs from the Working Paper's experimental LASSO approach against those from the more conventional and established FD24 approach.

4.3 RELIANCE ON A SINGLE MODEL VERSUS TRIANGULATION

124. Taking account of what is at stake in the base cost modelling, the Working Paper approach of placing 100% weight on a single water model and 100% weight on a single wastewater model is irrational. This is particularly so given that the two models used by the Working Paper have been selected using a novel algorithm-based approach. However, even if these models were of higher quality and had been subject to an effective development process, it would still not be reasonable to use only two models for the setting of these base cost allowances.

125. There are inherent difficulties in seeking to use econometric models to capture the influence of external factors on water companies' base costs, especially within a relatively small data sample that exhibits limited variation over time in most of the explanatory variables. We are also starting from a range of relevant cost drivers which when combined with a small sample size mean that a single cost model cannot estimate their impact accurately while including them all in a single model. In this context, we consider it essential that allowances are set using triangulation across multiple models, that differ in significant aspects of their design.

126. There is scope for reasonable debate as to whether, taking a step back, the FD24 model suite might be viewed as having too many models. But it is beyond any doubt that the Working Paper approach involves too few models.

127. Triangulation across multiple models, in the context of a small sample, provides a way to reduce the impact on allowances of the inevitable imperfections of any single model. It also reduces the extent to which the allowances for any one company are affected by specific modelling choices for which there is a lack of good evidence for relying entirely upon.

---

[66] Weeks Report, NWL-WPR-002, p.9.

128. Using triangulation is to show a sensible degree of humility in the face of econometric models that cannot hope to take account of the complex operational and business realities facing each water company. Reliance on a single model for the base cost modelling amounts to recklessness and indifference to the consequences of the modelling for companies and customers.

129. Professor Weeks criticises the lack of averaging/triangulation in the Working Paper approach:

> Model selection, the CMA's approach, is based on an untested sparsity assumption, and proceeds as if it is possible to identify the single correct specification and base all decisions on that specification alone. Any uncertainty about whether the chosen model is actually correct is effectively ignored once the selection is made.
>
> Model averaging, the approach traditionally used by Ofwat and Ofgem, explicitly acknowledges that no single specification can fully capture complex cost relationships. By combining predictions across multiple models, each capturing different aspects of cost variation, this approach hedges against any single misspecification. A company disadvantaged by one model specification may be better served by another; the averaged result provides a more balanced assessment.
>
> The CMA's exclusive reliance on LASSO-derived models deprives stakeholders of this insurance against model misspecification.[67]

130. There is evidence of the virtues of triangulation in the specific case of the water company base cost benchmarking in Section 6.2. We see that the FD24 model suites (involving triangulation across multiple models) performs well in terms of balancing goodness of fit and model stability compared to the individual performance of a very large number of feasible models.

[67] Weeks Report, NWL-WPR-002, p.iii.

5 THE WORKING PAPER MODELS

131. As set out in the preceding sections we have deep concerns about the process by which the CMA has produced, selected and reviewed its proposed base cost models. We also had material concerns with the quality of the models presented in the PD. While the Working Paper presents some changes to the approach and models since PD, fundamental concerns remain unresolved.

132. The Working Paper has not addressed the concern that **the selected models involve "over-fitting",** including too many explanatory variables within a relatively small data sample. This leads to the Working Paper models having multiple variables that would not be considered "statistically significant". And the allowances from the Working Paper models are much more unstable and sensitive to changes in the data than under the FD24 model suite.

133. Furthermore, the CMA has not adequately explained the large reductions in sector-wide allowances relative to the FD24 models (see Appendix 1). The reasons given by the CMA do not explain the extent of the difference, which is even greater than in the CMA's PD.

134. A summary of concerns raised in our Response to the PD on the quality of the models and whether they have been addressed in the Working Paper is provided in Figure 8.

FIGURE 8 WORKING PAPER RESPOSNE TO CONCERNS WITH THE QUALITY OF THE PD MODELS

Concern with PD	Action taken in Working Paper	Concern addressed?
1. Large reduction in industry-wide allowances vs FD24 that is not explained[68]	The Working Paper models provide bigger reductions to allowances, on average across companies, than in its PD. The CMA explains this by reference to energy prices and relative efficiency.	No – see Appendix 1.
2. Key modelling results much more sensitive / unstable to small changes in dataset[69]	The Working Paper makes some limited changes to its approach to select and review models	No – Results from Working Paper models still far less stable than under FD24 model suite. The Working Paper wrongly treats the sensitivity of the results to dropping companies as irrelevant.
3. Misrepresentation of statistical significance of the PD models due to failure to take account of panel data structure[70]	No change identified and issue not discussed in Working Paper	Working Paper ignores this issue and misrepresents the statistical significance of the explanatory variables
4. Lack of statistical significance of many of the variables in the PD	The Working Paper makes some limited changes to its approach to select and review models.	No – multiple variables in Working Paper models are not statistically significant at 5% and 10% thresholds

[68] NWL PD Response, section 2.1.4
[69] NWL PD Response, section 2.2.2
[70] NWL PD Response, section 2.2.1

Concern with PD	Action taken in Working Paper	Concern addressed?
models (especially when corrected for panel data structure)[71]		
5. Counter-intuitive model structure / modelling assumption in relation to impact of wages and energy prices on costs[72]	Working Paper revises the specification of energy and wage variables	Working Paper revised approach makes sense in principle (leaving aside wider concerns about number of variables and over-fitting)
6. Coefficients on wages and energy explanatory variables imply cost driver relationships that are not plausible[73]	Coefficients on wage and energy variables change as a result of change above (also impacted by wider changes in Working Paper LASSO implementation)	Coefficients seem to be too high, are unstable across different Working Paper models and are not statistically significant at 5% or 10% thresholds
7. PD Models difficult to interpret and sense check due to multiple density variables for which coefficients have different signs[74]	Working Paper reduces number of density variables in the selected models by using PCA but this reduces transparency and ability to sense check vs FD24 approach	Improvement from reduction in number of density variables, but the change of approach adds new interpretation issues and we have concerns with results under new PCA scale variables
8. Failure to use random effects for model estimation despite panel data structure and established practice[75]	No change identified and issue not discussed in Working Paper	No
9. Worse performance than FD24 models against Ofwat model robustness tests, especially in terms of misspecification, multicollinearity and sensitivity[76]	The Working Paper makes some limited changes to the approach to select and review models but problems remain	No
10. Insufficient consideration of data quality of explanatory variables used, especially APH variable[77]	Limited discussion of this issue in Working Paper	No
11. Use of a single model for each part of the value chain reduces overall reliability rather than triangulation across multiple models[78]	No change identified and no real explanation for Working Paper approach on this issue.	No - the Working Paper has used a single model for water and a single model for wastewater

135. In the remainder of this Section we elaborate on the following fundamental concerns:

- issues relating to overfitting and resultant unreliability of the Working Paper's modelling results (concerns 2, 3 and 4: Figure 8) (see Section 5.1); and
- concerns with coefficients on energy price variables (concern 6: Figure 8) (see Section 5.2).

[71] NWL PD Response, section 7.1.1
[72] NWL PD Response, section 2.2.3
[73] NWL PD Response, section 2.2.3
[74] NWL PD Response, section 2.2.3
[75] NWL PD Response, section 2.1.5
[76] NWL PD Response, section 7.1.1
[77] NWL PD Response, section 2.1.5
[78] NWL PD Response, section 2.1.5

136. We also discuss the serious concerns with data quality for the APH (see Section 5.3).

137. Ahead of addressing these issues, we comment briefly on the Working Paper's use of Principal Component Analysis (**PCA**) to derive variables to use in its modelling. The Working Paper considers that LASSO can be sensitive to multicollinearity amongst the set of candidate cost drivers and that PCA can help mitigate such problems. However, and within the context of the use of PCA, some implications and interpretation of the Working Paper's models raise serious concerns that the modelling has gone awry, namely with regards to how company scale, a critical cost driver, is being controlled for. A symptom of this is the finding in the Working Paper's water wholesale model that the smaller water only companies are disproportionately present in the set of companies deemed to be as or more efficient than the upper-quartile benchmark. Specifically, and in decreasing order of "efficiency", the set of those upper quartile companies are South Staffs, Portsmouth, Hafren Dyfrdwy, Affinity Water and Anglian Water. We are not confident that the CMA has given due consideration to the implications of the use of PCA. It appears that the CMA has adopted that approach with a view to addressing problems that the use of LASSO would otherwise have, as evidenced in the set of models produced at PD. As reflected elsewhere in our response, a more appropriate response would have been to move away from LASSO.

5.1 OVERFITTING AND UNRELIABILITY OF MODELLING RESULTS

138. The CMA states that its overriding objective was to "develop a simple suite of econometric benchmarking models designed to select economic and engineering cost drivers that most accurately predict routine, year-on-year costs, which companies incur in the normal running of the business".[79] Similarly the CMA states that "the primary test is whether the model, using cost drivers grounded in economic and engineering rationale, provides accurate predictions useful for economic benchmarking" and claims that "our updated models achieve this primary objective more effectively than Ofwat's models".[80]

139. However, the Working Paper adopts an over-simplified view of what it means for a modelling approach to provide "accurate predictions". The Working Paper's approach to both model selection and model assessment has given undue weight to goodness of fit over an historical period (e.g. RMSE). It has given insufficient attention to the stability of the estimation results and statistical significance which are relevant to predictability. In the context of a relatively small sample, this is an erroneous approach which a competent body would not be expected to retain for its final decision.

140. We consider this further in the context of:

- the approach to the analysis of the instability of model estimation results (see Section 5.1.1);

---

[79] Working Paper para 1.5.
[80] Working Paper para. 2.15.

- updated analysis of model stability between the FD24 and Working Paper models (see Section 5.1.2); and
- the statistical significance of model coefficients (see Section 5.1.3);

5.1.1 Approach to the analysis of the instability of model estimation results

141. A major concern that we found with the PD models was that these were more sensitive – and less stable – to small changes in the dataset from dropping years and/or companies.

142. This type of analysis is an important part of the model assessment and selection process, especially in the context of models generated by a novel LASSO process that have more explanatory variables and may be subject to over-fitting.

143. The Working Paper includes some new analysis, not covered in the PD, which involves a form of bootstrap resampling which the CMA says allows it to "observe the stability of the model's predictions under different data realisations".[81]

144. There is some common ground between aspects of the CMA's bootstrapping approach to model robustness and the sensitivity / stability analysis we carried out for our PD Response. We consider that looking at the stability of predicted costs to variations in the dataset can provide valuable evidence on the performance or robustness of a model or set of models. While we used the terminology of the "sensitivity" of modelled costs in our PD Response, our assessment was absolutely one of assessing "the stability of the model's predictions under different data realisations".[82]

145. There are however important differences between our approach and the CMA's in relation to:

- how the results from the stability analysis are used; and
- how stability is assessed.

146. On the first point, the Working Paper's exercise involves no like-for-like comparison with the FD24 models. The Working Paper declares that the results of its models are robust by reference to the position of the median value in the distribution of post-efficiency cost allowances for the Disputing Companies and the spread of these values.[83] However, this exercise is entirely focused on the Working Paper's models/approach. For instance, the Working Paper states that the interquartile range is "narrow" for the water and wastewater models and in turn that the results are robust.[84] However, given the CMA accepts that the stability or predicted value is an important aspect of model assessment, we see no justification for there being no comparative assessment of whether

81 Working Paper, para 2.27.
82 Working Paper, para 2.27.
83 Working Paper, para 3.34.
84 Working Paper, paras 3.34 and 4.32.

the distribution is wider or narrow under the Working Paper's models than under the Fd24 models. This omission is especially concerning since we had presented evidence that the results from the PD models were less stable to dataset variations than those from the FD24 models.[85]

147. On the second point above, the CMA wrongly dismisses the type of analysis that we carried out in our PD Response to look at the stability of modelled costs.

148. Our approach of dropping whole companies and/or whole years from the dataset is an established technique in the context of panel data. It is a form of "Leave-One-Group-Out (**LOGO**) resampling". This has some similarities with the Leave-Out-One (**LOO**) resampling approach that the CMA has used for its LASSO cross validation (**LOOCV**), but rather than leaving out just a single observation at a time, we leave out groups (companies and/or years). This is a more powerful approach to assessing robustness in a context where there is panel data rather than a dataset of independent observations. It respects the panel structure of the dataset, particularly when much of the variation in the data is between companies rather than between observations for a single company.

149. The Working Paper says that that tests involving dropping an entire company from the sample are not useful for checking the appropriateness of a benchmarking model.[86] The Working Paper provides no cogent explanation for its position and it is essential for the CMA to revisit its position on this matter.

150. One argument made by the Working Paper is that a test cannot be selective in which parts of the data are held out of the sample. We do not think this issue relates to the analysis we provided, which is systematic: the resampling analysis involves all companies being dropped in turn and no company being given more / less weight than others. This argument might apply to evidence from other parties which differs from that which we are concerned with here.

151. The other argument made by the Working Paper refers to the special water regime which it says "may in fact prohibit an equivalent loss of company data from that available for benchmarking in the event of a merger".[87] The reference to the special water regime may reflect comments from Ofwat that "in recent mergers, we and the CMA have found that the loss of a comparator negatively impacts the models" and that "[i]t would not be consistent to require the CMA to produce models which are robust to that finding".[88] These arguments about the special merger regime from the CMA and Ofwat are entirely misplaced: they are simply not relevant to the task at hand of comparing the robustness of alternative models for the purposes of setting base cost allowances at PR24.

---

[85] NWL PD Response, Figure 37.
[86] Working Paper, para 2.29.
[87] Working Paper, para 2.29.
[88] Ofwat, Response to Base Cost Modelling, 18 November 2025, para 1.9.

152. Professor Weeks' report agrees with our positioning on the relevance of leave out company out checks when he says:

> The CMA's statement that tests "in which an entire company had been removed from the sample" are not "useful for checking the appropriateness of a benchmarking model" appears to draw on merger-regime reasoning inappropriately applied to model validation. The merger context asks: "Does losing Bristol Water permanently from the comparator set harm Ofwat's regulatory function?"
>
> In contrast LOCO diagnostics ask: "Does the model generalise to each individual company, or is its apparent accuracy an artefact of having seen that company's data?" These are fundamentally different questions. The merger analysis concerns sample adequacy for long-term regulatory purposes; LOCO concerns whether the specific model specification captures genuine cost relationships that apply across all companies.[89]

153. To be clear, under our approach to the analysis of model stability we are not expecting the results from a specific model to be entirely stable or invariant to the loss of a company from the dataset. Instead, our position is that if the results from model A are *more unstable* than the results from model B, under the exercise of systematically dropping companies from the sample, this is a significant disadvantage of model A compared to model B. Stability is not the only relevant dimension of model performance, but it is an important one. This is because it shows how much the observed relationship is driven by a single company versus data from all companies – where a model's results are dependent on the data from a single company this is a strong sign that the relationship is not supported by the remainder of the data and that other models may be better. The less stable the results are the more likely it is that they are driven by spurious or random correlations in the data than by the model capturing underlying causal relationships between costs and cost driver variables.

154. It may very well be the case that the loss of a company from the dataset is prevented by the special water regime, but that does not detract in any way from the value of the analytical exercise of looking at how alternative models compare in terms of the impacts on results from dropping companies. Looking at the stability of benchmarking results to dropping companies is something that has been done for many years by practitioners working on water company econometric benchmarking.

155. This has also been Ofwat's established approach, including throughout PR24. For instance, Ofwat's own approach to model robustness at PR24 explicitly considers the impacts of dropping whole companies ("Sensitivity of estimated coefficients to removal of most and least efficient company").[90] Our approach to stability builds on Ofwat's approach of dropping some companies, but applies it in a more systematic way (dropping all companies in turn rather than the most/least "efficient") and in a more targeted way (looking directly at the impact of dataset variations on

---

[89] Weeks Report, NWL-WPR-002, p.6-7.
[90] Ofwat FD24, PR24 final determinations Expenditure allowances – base cost modelling decision appendix, p. 63.

modelled costs).

156. The CMA has applied a form of bootstrap resampling for its analysis of model stability that does not involve dropping / replacing whole companies or years from the dataset. There may be versions of bootstrap approaches that are better suited to the panel data structure, but the approach used for the Working Paper seems less appropriate to the dataset under consideration, and the purposes of the base cost modelling, than the resampling approach that we have used.

5.1.2 Updated analysis of model stability between Ofwat and CMA models

157. As discussed above, the Working Paper does not provide any comparative analysis of the stability of modelled costs under the Working Paper models versus the FD24 model suites (or other potential models). While there are differences in the approach to resampling we consider that our approach is valid and enables such comparisons to be made in practice.[91]

158. Figure 9 presents a version of the Working Paper Figure 3.3 which we have adapted to present results from our approach to the analysis model stability as described in the subsection above (and as used in our PD response). It allows for comparisons between the Working Paper's wholesale water models (from its working paper) and the FD24 wholesale water model suite.[92] Whereas we presented histograms in our PD response,[93] the same analytical results can be used to produce the type of box plot presented by the CMA.

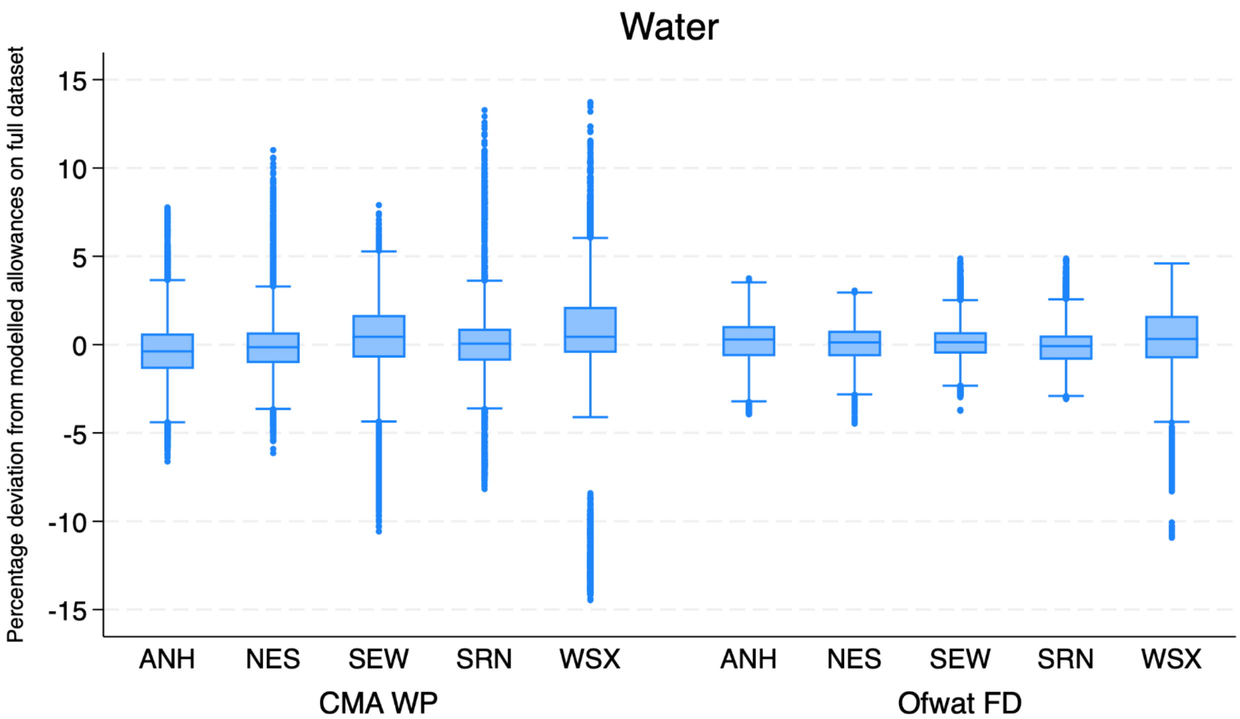
159. For the wholesale water modelling, we see that the spread of modelled costs is considerably narrower under the FD24 models, for each of the five companies. The CMA clearly assessed a narrower spread of the distribution as a desirable feature; on that basis the FD24 model suite is better than the Working Paper's model in terms of stability.

[91] Our approach, based on dropping all combinations of companies and years, one combination at a time (i.e. dropping observations for company X and observations for year T, for each pairing of year-company), allows for an assessment of the stability of Ofwat's model suite (which involves averaging/triangulation across a set of models), taking account of how diversity across these models can help the triangulated modelled costs be more stable than the modelled costs for individual models. This can be compared against the CMA's models (or indeed a suite of CMA models if one were to exist). Without an assessment at the suite level, the comparison between individual Ofwat models and the CMA's approach would be misleading. The CMA's omission of any comparison of stability with Ofwat's model suite may reflect challenges of assessing stability for a suite of models using standard software code. We can provide further guidance to the CMA on our computational code if needed.

[92] The chart is a boxplot of the percentage difference between (i) companies' modelled costs in each year over the period 2012 to 2024 when these are derived from regressions using a variant of the dataset, and (ii) their modelled costs based on the unchanged dataset. The boxplot shows the variation in that percentage over the variants of the dataset outlined in footnote 93

[93] NWL PD Response, Figure 8.

FIGURE 9: COMPARISON OF STABILITY OF MODELLED COSTS FOR WORKING PAPER VS FD24 MODELS



Source: NWL analysis. NWL WPR Databook.

160. It can be difficult to make comparisons purely from charts such as that above. For that reason, we use an instability metric, which we calculate as the standard deviation from the distributions shown in the chart above, taken across all companies.[94] The standard deviation is good choice for the metric as it takes direct account of all of the observations (the inter-quartile range ignores not just extreme observations but just under half of the observations). We show results on this basis in Figure 10. This metric is the same as that which we used in our PD response.

FIGURE 10: COMPARISON OF MODELLED COSTS INSTABILITY METRIC (STANDARD DEVIATION OF MODELLED COST DEVIATIONS) BETWEEN WORKING PAPER AND FD24 MODELS

	Water	Wastewater
Working Paper model	4.62	3.49
FD24 model suite	2.02	1.87

Source: NWL analysis. NWL WPR Databooks.

161. This analysis shows that the Working Paper models are significantly less stable than the FD24 model suite.

162. While the Working Paper presents analysis of the purported stability of the Working Paper models (or LASSO process) our analysis indicates that the FD24 model suite performs better in terms of stability. This is partly driven by the greater stability (i.e. lower instability) of the individual FD24

94 NWL PD Response, Annex 1, Section 7.2.1.

models (e.g. from fewer explanatory variables, given the small sample size which allows coefficients to be estimated more accurately) and partly from the benefits of triangulating over multiple models.

163. Further analysis of the relative performance of the Working Paper's model in terms of both stability and goodness of fit is provided in Section 6.2.

5.1.3 Statistical significance of model coefficients

164. The Working Paper pays no regard to the statistical significance or standard error of the coefficients estimated for the models selected by its experimental LASSO approach. This is a remarkable position for the CMA to adopt.

165. As set out in the tables below, the Working Paper's models include multiple variables for which the estimated coefficients would not be considered statistically significant from zero at the 5% and even 10% thresholds. The Working Paper's models are not fit for purpose. This is further evidence that the specific LASSO implementation that the Working Paper adopts has failed to guard against over-fitting.

166. The Working Paper *misrepresents* the statistical significance of the coefficients for the explanatory variables in its models. The Working Paper reports figures for statistical significance that would be valid if the data sample contained a set of independent observations. It is not valid for a data sample containing annual data for a set of companies over a 13-year period. Professor Weeks reaches the same conclusion.[95]

167. This is a technical error, which has the tendency to over-state the statistical significance of the variables in the models. This error was raised in responses to the PD.[96] The CMA's failure to correct it indicates problems with the CMA's internal quality assurance. It is alarming to see the CMA ignore the relevance of the panel data structure of the data in reporting statistical significance.

168. This is an area where there has in the past been a large degree of common ground between Ofwat, the water companies and the CMA. We have used the established practice of cluster robust standard errors in Figure 11 and Figure 12 to assess the Working Paper models.

169. Combined with the complementary evidence from the resampling analysis, the Working Paper models do not seem fit for purpose.

[95] Weeks Report, NWL-WPR-002, pp. 7-8 on "Uncertainty from Incorrect Standard errors".
[96] NWL PD Response, para 348.

FIGURE 11: STATISTICAL PERFORMANCE OF THE WORKING PAPER WATER MODELS – OLS WITH CLUSTER ROBUST STANDARD ERRORS.

Cost driver	Explanatory variable	Water Resources Plus model (WRP)	Treated Water Distribution model (TWD)	Wholesale Water model (WW)
Scale	Properties (log)	1.049***		
	Length of mains (log)		1.046***	
	Scale combined (log)			-0.683***
Complexity	Weighted average treatment complexity (log)	0.674**		0.450***
Topography	Average pumping head TWD (log)		0.296***	0.096
	Booster pumping stations per length of mains (log)		0.338**	0.318**
Density	Density combined (log)	-0.091*	0.154***	0.050***
	Density combined (log) - squared	0.017	0.039***	0.026***
Input price	Energy index (log)	0.179	0.127	0.116*
	Construction wages (log)	-0.217	0.556	0.303
Constant	Constant	-11.734***	-7.666***	3.804***
Robustness tests	Adjusted R-squared	0.910	0.969	0.971
	RESET test	0.550	0.765	0.065
	VIF	2.074	2.329	2.4
	Normality	0.215	0.394	0.139
	Heteroskedasticity	0	0.19	0

Source: regression output of the CMA water models using cluster robust standard errors

FIGURE 12 : STATISTICAL PERFORMANCE OF THE WORKING PAPER WASTEWATER MODELS - OLS WITH CLUSTER ROBUST STANDARD ERROR.

Cost driver	Explanatory variable	Sewage Collection model (SWC)	Sewage Treatment model (STW)	Wastewater Network Plus model (WWNP)
Scale	Load (log)		0.842***	
	Sewer length (log)	0.825***		
	Scale combined (log)			-0.307***
Economies of scale	Weighted average treatment size (log)		-0.215***	-0.018
	Load treated in size bands 1 to 3 (%)			0.004
Complexity	Load treated with ammonia consent <=3mg/l (%)		0.004	0.006***
Topography	Pumping capacity per sewer length (log)	0.431***		0.467***
Density	Density combined (log)	0.099***		
Urban rainfall	Urban rainfall per sewer length (log)	0.118**		0.115**
Input price	Energy index (log)	0.137	0.292***	0.167
	Construction wages (log)		0.094	
Constant	Constant	-4.467***	-5.387**	5.013***

Robustness tests	Adjusted R-squared	0.916	0.906	0.95
	RESET test	0.025	0.062	0.194
	VIF	2.159	5.015	5.071
	Normality	0.102	0.023	0.065
	Heteroskedasticity	0.015	0.724	0.033

Source: regression output of the CMA's wastewater models using cluster robust standard errors

5.2 CONCERNS WITH COEFFICIENTS ON ENERGY PRICE VARIABLES

170. The Working Paper changes the way that the energy price variables are structured.[97] This tackles the specific concern with the PD modelling that it did not make sense to multiply energy prices by a measure of scale (e.g. mains length) before including it as an explanatory variable in the models.

171. However, there remain severe concerns about seeking to take account of energy prices through econometric models given limitations in the accuracy with which coefficients are estimated, especially given small sample sizes and other changes over time that are not captured in the Working Paper modelling.

172. The Working Paper states that the coefficients on the energy price variable in its water model has the "expected sign" and is of "plausible magnitude".[98] The Working Paper does not explain why the CMA considers the coefficient on the energy price to be plausible or provide any evidence to support its assessment. For wastewater the Working Paper does not comment directly on the corresponding coefficient.

173. Our assessment is that the coefficients on the energy price variables are considerably higher than the level that can be explained by energy prices alone, indicating that they are picking up other factors besides energy. This is supported by the assessment undertaken by Economic Insight.[99]

174. This issue is more complicated than it might first appear because of the interactions with the CPIH. The CPIH rose considerably over the period covered by the data sample, and one important driver of this was energy costs. CPIH inflation has already been stripped out of the input data used for the modelling; all expenditure data has been adjusted to 2022/23 price base using the CPIH.

175. Given this, it is far from obvious what the energy coefficient in the CMA's models is meant to be picking up and how, in turn, we can assess whether the coefficient is plausible.

176. Our assessment is as follows in relation to the coefficient on the logarithm of CPIH-adjusted energy prices where the dependent variable is CPIH-adjusted base expenditure:

- for a company that has a very lower power share of costs we would expect a negative coefficient for the relationship between real energy prices and real expenditure as its costs

[97] Working Paper, para 2.12.
[98] Working Paper, para 3.18.
[99] EI Energy Report, NWL-WPR-001, Section 2, p.6-7.

would increase by less than CPIH during high energy price inflation;

- for a company with a power share similar to the average power share across goods and services feeding into the CPIH, we would expect the coefficient to be around zero as its overall costs would rise in line with CPIH; and
- for a company with a greater power share of costs we would expect a positive coefficient but it would be considerably less than the power cost share, as during a period of high energy price inflation we would expect its costs to outstrip CPIH growth.

177. We recognise that this is a relatively complicated issue. We provide simplified bottom-up spreadsheet modelling to help illustrate and demonstrate the critical points about interactions with CPIH.[100]

178. Turning back to the case at hand, two pieces of evidence seem to be particularly informative:

- across the data period used for model estimation from 2011/12 to 2023/24, the average share of power costs to modelled base costs is 12.1% for wholesale water; and 13.4% for wastewater network plus; and
- the ONS has estimated that the Consumer Prices Index (**CPI**) has an indirect energy intensity (the proportion of the consumer price of products accounted for by energy costs) of 6.6%.[101]

179. On this basis, we would expect a positive coefficient for water companies given that power costs shares are above those in the industry as a whole.

180. We think an approximation for the expected coefficient (elasticity) for the energy price variable in the Working Paper models can be obtained by deducting the estimate of the energy intensity in the CPIH (6.6%) from the power cost share. We found through our illustrative bottom-up spreadsheet modelling, across a range of scenarios, that such a deduction provides a more accurate prediction of the coefficient (elasticity) than assuming that the coefficient should equal the power cost share.[102]

181. On this basis, we might expect the coefficient for the impact of energy prices on base costs (CPIH-adjusted) to be approximately 5.5% for water (i.e. 12.1% minus 6.6%) and 6.8% (i.e. 13.4% minus 6.6%) for wastewater. These are approximate figures, but a more reliable and intuitive guide than using the energy share alone and wrongly ignoring the deflation of costs and input prices by CPIH.

182. We list the elasticity of energy price to CPIH-adjusted costs that is implied by the Working Paper's

[100] See NWL WPR Databook "Illustrative energy price - CPIH interactions".
[101] The energy intensity of the Consumer Prices Index: 2022
[102] See NWL WPR Databook, "Illustrative energy price - CPIH interactions".

models in Figure 13. We can see that these are significantly above the levels we would expect.

FIGURE 13: ELASTICITIES OF ENERGY PRICE TO CPIH-ADJUSTED COSTS IMPLIED BY WORKING PAPER MODELS

Model	Elasticity implied by coefficient	Comment on magnitude	Statistical significance
Wholesale water			
WP top-down	11.6%	Coefficient too high	Not statistically significant at 5% level
WP TWD	12.7%	Coefficient too high	Not statistically significant at 10% level
WP WRP	17.9%	Coefficient too high	Not statistically significant at 10% level
Wastewater network plus			
WP top-down	16.7%	Coefficient too high	Not statistically significant at 10% level
WP SWT	29.2%	Coefficient too high	Statistically significant at 1% level
WP STC	13.8%	Coefficient too high	Not statistically significant at 10% level

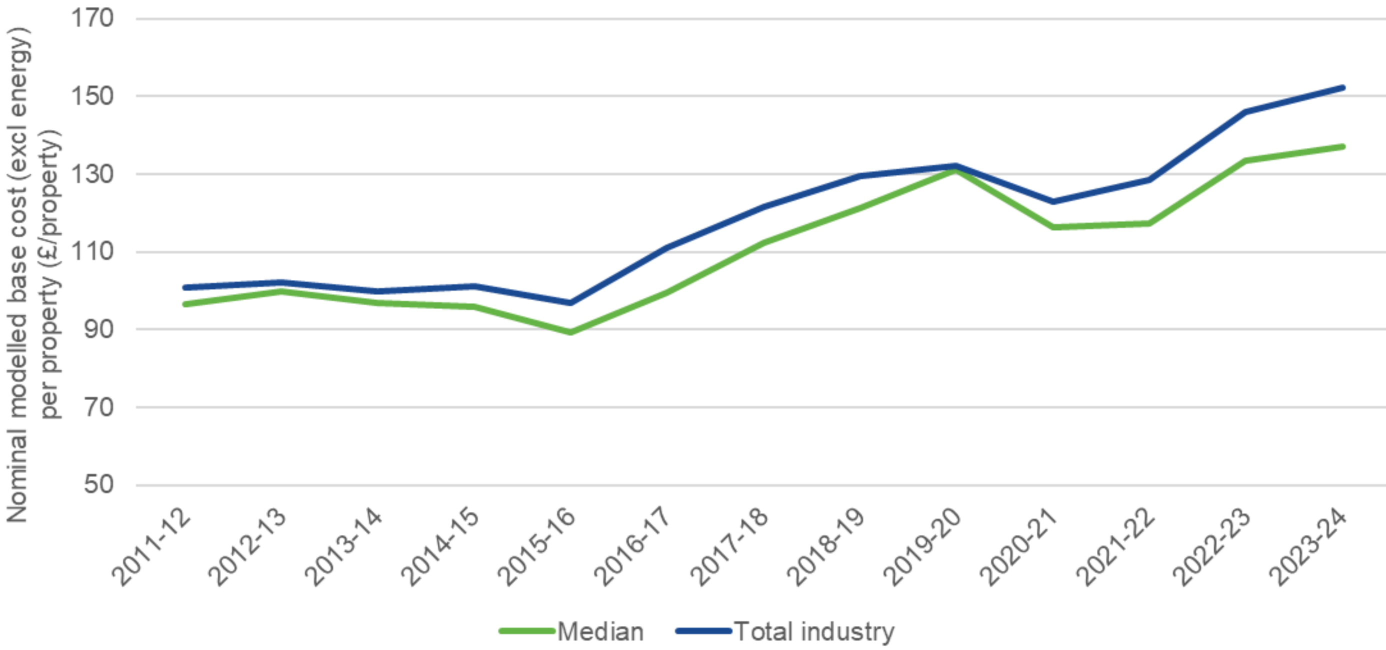
Source: coefficients of the power index variable from all CMA Working Paper models

183. Further to these issues, the Working Paper's modelling approach suffers from the attempt to fit a one-size-fits-all coefficient to the elasticity of costs to energy prices. We would expect, and we can see from our worked example spreadsheet, that the expected coefficient would vary across companies and over time as the power share for individual companies and within the CPIH changes.

184. We consider that the energy price coefficients from the Working Paper models, including the two top-down models selected in the Working Paper, are misleading and are capturing effects other than energy price effects. Since real energy prices have increased substantially over the sample period the coefficient on the energy price variable will tend to be artificially inflated by any increases over time in industry-wide base costs that are not due to energy prices and not captured by the other explanatory variables in the model.

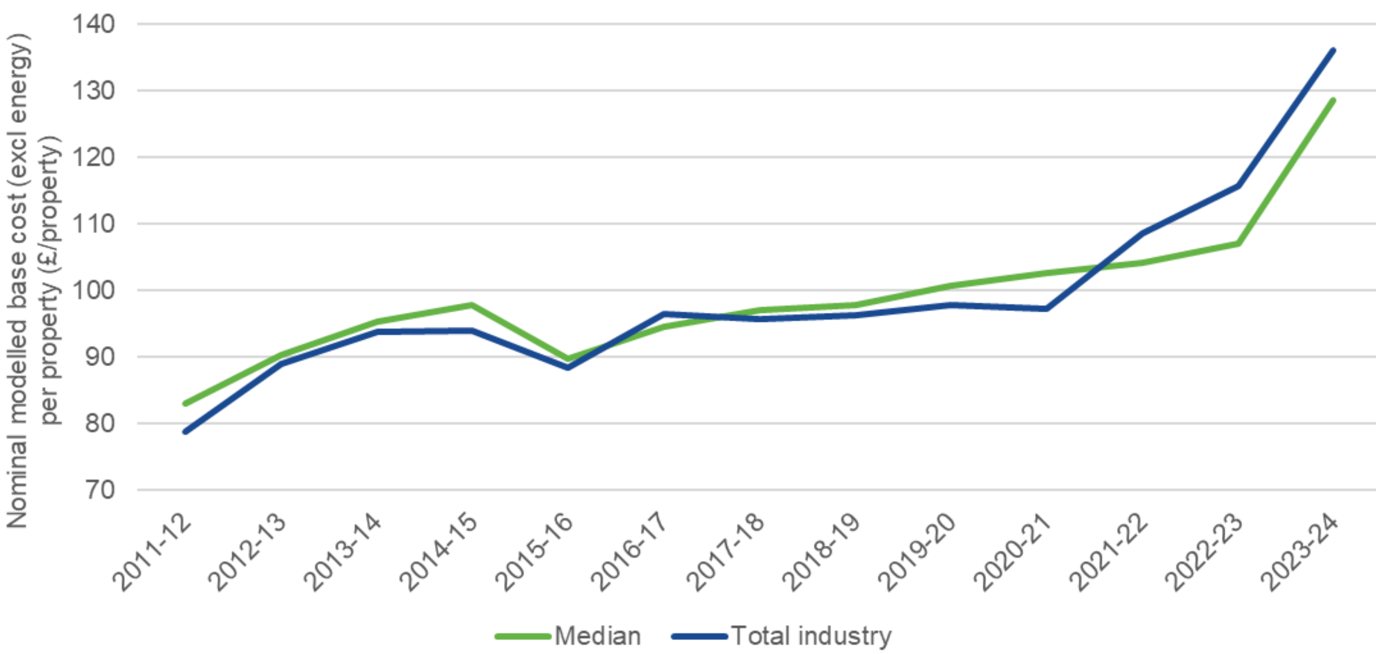
185. Figure 14 and Figure 15 below show increases in costs per property supplied over time excluding power costs. We see that there have been significant real-terms increases in costs per property over time that are unrelated to energy costs. These increases have distorted the Working Paper's modelling of energy prices.

FIGURE 14: BASE MODELLED COST (EXCLUDING ENERGY) PER PROPERTY – WATER



Source: NWL analysis of base modelled cost data from Ofwat's feeder model 1 . NWL WPR Databook

FIGURE 15: BASE MODELLED COST (EXCLUDING ENERGY) PER PROPERTY - WASTEWATER NETWORK+



Source: NWL analysis of base modelled cost data from Ofwat's feeder model 1. NWL WPR Databook.

186. There is further evidence that the Working Paper modelling is wrongly attributing the drivers of cost increases over time to energy prices. Following what we did at PD, we carried out analysis of what happens under the Working Paper's LASSO process if we entirely exclude power costs

from the dependent variable and re-run the model selection and estimation process. We found:

- the energy prices explanatory variable for are selected for both water and wastewater top-down models;
- the coefficient for the water top-down model is positive at 0.043; and
- the coefficient for the wastewater network top-down model is positive, at 0.038.

187. The first point above is further evidence of the deficiencies of the LASSO process. The LASSO process is selecting variables that happen to fit the data; it is failing to filter out variables that do not have any material causal relationship with costs.

188. The results for the two coefficients above reinforce the view that the coefficients from the Working Paper models are capturing changes over time in base costs that are not related to energy prices. There seems to be a positive correlation between energy prices and base costs excluding power, but this is not a relationship that makes any sense.

189. It is deeply alarming that the CMA is proposing to introduce energy price variables into the base cost models for the first time, despite its selected models having coefficients on the energy price variables that are not statistically significant at the 5% level, which seem far higher than can be explained by a proper economic analysis of the evidence and which are clearly capturing effects other than energy prices.

190. There is additional evidence on the problems with the inclusion of energy price variables in the modelling in the Economic Insight Report.[103]

5.3 THE NEED TO REVISIT THE INCLUSION OF AVERAGE PUMPING HEAD IN THE WATER MODELS

191. We do not understand the rationale for the position that CMA has adopted on APH.

192. In the PD the CMA noted the concerns raised by Southern regarding APH and data quality but did not discuss this issue at all in its own assessment. The CMA said:

> 4.35 We considered that assessing these different submissions in a piecemeal way would be neither practical nor conceptually correct. This is because the contribution of a variable to the explanatory power of a model depends not just on the economic rationale for its inclusion or the quality of its measurement, but also on its relationship with the other variables included.
>
> 4.36 Instead, we have provisionally decided to assess these eight different issues under a unified framework, using an econometric approach known as LASSO.[104]

193. This approach avoids the issue. LASSO cannot assess whether a variable is measured accurately

[103] EI Energy Report, NWL-WPR-001, Section 2.
[104] PD, p. 47.

- all it measures is how well it improves goodness of fit. It is quite possible that a variable may well be correlated with costs but still be poorly measured to the extent that it provides spurious implied relative efficiencies – this was the view reached by the CMA at PR19. We therefore do not consider that the CMA has given due consideration to the issues raised around data quality.

194. In the Working Paper the CMA states:

> To control for differences in cost due to network topology, we include average pumping head (APH) in our updated bottom-up TWD model and our top-down model. Since our PR24 PD, we have not seen sufficient evidence to change our view on the data quality of the APH cost driver, so include this in our updated bottom-up TWD model and our top-down wholesale model.[105]

195. Given that the PD contains no indication of a detailed assessment of the APH data quality challenge, we are unclear what analysis and reasoning the CMA conducted in order to reach this conclusion.

196. We ask the CMA to reconsider its position in this area as it is a material issue for us and other appellants. At the very least it should set out its full analysis and reasoning in the FD to provide transparency on why it considers that the concerns raised around APH data quality can be dismissed. In particular, we think there are two specific issues that the CMA must address:

- how can improved APH data since the Turner and Townsend report have been accurately back-casted to the historical period in the data sample?; and
- how is the CMA comfortable about the level of estimated rather than measured data to estimate APH and the reliability of this data?

197. We discuss our views on these issues in turn.

5.3.1 The accuracy of backcasted data

198. The data relied upon by Ofwat and the CMA for its base cost models is set out in "base costs – water model 1".[106] Within this sheet Ofwat has modified historical data for treated water distribution APH (the variable used by Ofwat and the CMA) for 9 of the 17 companies in the "Input_overrides" tab. In total, 76 of the 221 data points have been modified which represents over a third of the dataset. The "Override explanation" sheet provides limited explanation of the changes. For example:

- the changes for Anglian are predominantly explained by "Adj. based on adjustment applied by ANH in 20-21 and 21-22";
- for United Utilities, Thames and Wessex, it just says "APR query response"; and

---

[105] Working Paper, para 3.7.
[106] Ofwat PR24 FD CA03 Base costs water model 1.

- for us and Southern the explanations are very similar. Ours says "NES average pumping data presents significant volatility across levels of aggregation over time (ie significant reallocations from one level of aggregation to another over years). To correct for this, we re-allocate total APH in each year across levels of aggregations, using as weights the average weight of each level of aggregation in 2020-21 and 2021-22 (since these two years should be better quality). This means that total APH in each year will be unchanged, but data in each level of aggregation will present less volatility over time.".

199. The changes made by these overrides are not immaterial. Figure 16 below shows significant changes to 2011-12 data for 8 companies.

FIGURE 16: CHANGES TO COMPANIES' 2011-12 AVERAGE PUMPING HEAD (DISTRIBUTION) DATA

Company	2011-12 original input	2011-12 overwritten input	% change
ANH	74.23	68.86	-7%
NES	20.50	56.63	176%
NWT	45.43	65.70	45%
SRN	103.02	68.21	-34%
TMS	65.20	50.14	-23%
WSX	18.17	73.03	302%
YKY	97.10	75.61	-22%
SES	100.78	89.12	-12%

Source: NWL analysis of Ofwat FD base cost water dataset (model 1). NWL WPR Databook.

200. We find this approach to rewriting history with such limited explanation and justification astonishing. It also flies in the face of the Turner and Townsend report that Ofwat commissioned in the run up to PR24. That report set out a number of areas of concern that Ofwat would need to consider if using APH data for cost assessment. One of these was that:

> If companies move to using more measured data instead of estimating it is unlikely they will be able to accurately back-cast the measured data. The relationship between measured and estimated data may not be consistent over time.[107]

201. Given this conclusion from the most detailed review of APH data by experts in the field for the economic regulator itself, we struggle to see how Ofwat and the CMA can be confident that this backcasted data is accurate.

5.3.2 The level of estimated data

202. One of the problems with APH data is that it has been based on estimated rather than measured data, which has largely fed the concerns around its data quality. This has improved in recent years but is still not convincing.

203. Figure 9 of Ofwat response to the companies' SoCs showed that the percentage of APH data

[107] Turner & Townsend and WRc, Average Pumping Head - Data Quality Improvement p. 46.

from measured data has increased from 60% to 72% between the 2022 Turner and Townsend report and the 2024 update.[108] Whilst an improvement there is still cause for concern:

- this still means that 28% of the data is not based on measured data at all. The same cannot be said for any other variable used in the base cost modelling. In particular, there are no such concerns about the number of booster stations which is the alternative to APH that the CMA considered to be a "superior alternative" in the PR19 appeal;[109] and
- some companies still have some very low levels of measured data being used to calculate APH.[110] For example, United Utilities which is the 3rd largest company in the sector by base costs, only has 5% of its data coming from measured data. Wessex, another WaSC, only has 26% of its data from measured data. This does not seem a reliable basis for using within cost models worth billions of pounds to companies and customers.

5.3.3 Suggested steps on APH

204. We are also not surprised that the APH variable used in the Working Paper's top down water model is not statistically significant when appropriate robust standard errors are used (see Figure 11). This is consistent with the concerns set out around the data quality of the variable.

205. To ensure that this is properly taken account of we encourage the CMA to:

- **ensure that it has properly considered the data quality of APH given the concerns raised**: This is not a task that can be delegated to the LASSO algorithm. If the CMA decides that the variable does have good quality data, the CMA must share that analysis and evidence to support this position in its final determination;
- **remove the APH variable from the modelling**: we do not see how the concerns over the APH variable have been resolved. The backcasted data cannot be accurate and the current data still relies heavily on estimated data, and very strongly for individual companies. This is likely to be distorting the results and giving a misleading picture of efficient costs for different companies; and
- **mitigate the risks through averaging**: at the very least the CMA should adopt the Ofwat approach of using models that include the APH variable (but not the booster stations variable) and models with the booster stations variable (but not the APH variable), and then average between them to ensure that the concerns around data quality are partially acknowledged. The inclusion of both variables does not achieve this as they cannot be estimated accurately when used in tandem and the unreliable APH data will be unduly influencing the single model being used.

[108] Ofwat, PR24 Redeterminations – Expenditure allowances – common issues, Figure 9, p. 35.
[109] CMA PR19 Final Determination para 4.83.
[110] Ofwat databook to support PR24 Redeterminations – Expenditure allowances – common issues, Figure 9 (see FN 108).

6 HOW THE CMA CAN SELECT MODELS FOR ITS FINAL DETERMINATIONS

206. The Working Paper is highly problematic. It fails to address the fundamental problems with the approach taken in the PD and presents an implausible and unjustified reduction in base cost allowances. We see no valid way for the CMA to use the Working Paper models as the basis for setting allowances for AMP8.

207. Taking account of the limited time left for the CMA's redetermination, we have identified some practical actions for the CMA to take in moving from the Working Paper to its final determinations. The steps and options outlined below apply to model selection and the treatment of energy prices. They are additional to our conclusion that the upper quartile "catch-up" adjustment is no longer a valid regulatory adjustment to apply when setting allowances (see Section 3).

208. In particular we consider:

- the appropriate starting point for the CMA's final determination (see Section 6.1);
- the approach to model selection and triangulation (see Section 6.2); and
- potential refinements to Ofwat's energy price adjustment (see Section 6.3).

6.1 THE STARTING POINT FOR A BETTER APPROACH FOR THE CMA'S FINAL DETERMINATION

209. The three problems that seem most important to address are as follows: (a) the Working Paper's models are unstable and unreliable due to overfitting; (b) the Working Paper's models include coefficients for energy prices that are excessive and are likely to be picking up other factors besides energy prices thereby distorting the overall modelling; and (c) the Working Paper's approach is severely compromised by seeking to rely on a single econometric model for each of water and wastewater.

210. To address this, there is a need for the CMA to use:

- models that have more stable results and statistically significant explanatory variables;
- triangulation across models that involve different variables and approaches; and
- an alternative more reliable approach to allowing for the impacts of energy price changes.

211. In addition to these overarching requirements, there is a need for the CMA to avoid undue reliance on a model (or models) that include the average pumping head variable, due to the data quality issues discussed in Section 5.3.

212. Given the late stage in the CMA process, and the needs above, we see a strong case for the CMA reverting to the FD24 models for its PR24 redetermination. The FD24 model suite strikes a far better balance between goodness of fit and the stability of results than the Working Paper's proposed models. The FD24 model suite is more stable and benefits from triangulation across a range of individual models that differ in significant ways. It has also been subject to extensive

consultation and development over time.

213. Setting base cost allowances by giving 100% weight to the figures from the Working Paper does not make rational sense given the way the figures have been generated and the context in which allowances are to be set: large over-spends across the industry in AMP7 (that are mostly unrelated to energy prices), and major asset health and performance challenges in AMP8.

214. In relation to the treatment of energy prices, we recognise that Ofwat's approach at PR24 has some shortcomings. We discuss in Section 6.3 how the CMA could make targeted changes to Ofwat's post modelling adjustments. We also note that these adjustments draw on similar calculation logic and data to that envisaged by the CMA under its energy price true-up proposals.

215. The three elements above are each needed in their own right and it would not be adequate for the CMA to respond to only one or two of these**.** For example, it would not be sufficient to simply triangulate across a set of LASSO-derived models that suffer from the over-fitting and/or energy price concerns (e.g. across the bottom-up and top-down LASSO models), or across a set of models that come from minor changes to the way that the CMA implements LASSO). Similarly, it would not be sufficient to remove energy price variables from the LASSO process while retaining the rest of the LASSO approach.

6.2 MODEL SELECTION AND TRIANGULATION

216. We have taken a step back and considered how the CMA might approach model selection for its final determination, given the clear deficiencies of its LASSO approach and its Working Paper models.

6.2.1 Looking at the trade-off between RMSE and stability across a fuller set of candidate models

217. The potential value of techniques such as LASSO tends to come in situations where the datasets, and in particular the number of candidate explanatory variables, are far larger than that used for water company base cost benchmarking.

218. We found that, starting with the set of candidate explanatory variables used in the Working Paper modelling, it is possible to estimate and compare *all possible* models that involve these variables.[111] For example, for the set of eight candidate explanatory variables used for the Working Paper's top-down water modelling there are 255 possible models.[112] This number of combinations is sufficiently low that, with modern computer processing speed, it is possible to

[111] Given the small sample, and in line with the Working Paper and the FD24 approach, we do not use interaction terms for this exercise.
[112] The more general formula is that if there are n candidate explanatory variables there are 2^n -1 possible models (leaving aside a model with no candidate variables). Given this formula there is an exponential rise in the number of feasible models with the number of variables under consideration: 10 variables leads to around 1000 models while 50 variables leads to over a quadrillion models.

carry out a comparative assessment of all of these candidate models.

219. In Figure 17 we compare each of the 255 possible top-down wholesale water models along two key dimensions of model performance: (a) goodness of fit, measured by RMSE (the metric used extensively by the CMA); and (b) the model instability metric we described in Section 5.1.2 above (which captures the extent to which modelled costs over the sample of companies and years vary under changes to the dataset from dropping a single year and/or company).

220. In Figure 17:

- the red dot shows results for the selected top-down model from the Working Paper;
- the yellow dot represents the FD24 model suite;[113]
- the blue dots represent other candidate models that represent frontier positions in the sense that there are no other models with better performance on the RMSE and instability metrics. The curve of frontier models reflects the trade-off between goodness of fit and stability; and
- the grey dots show other feasible models not covered above.

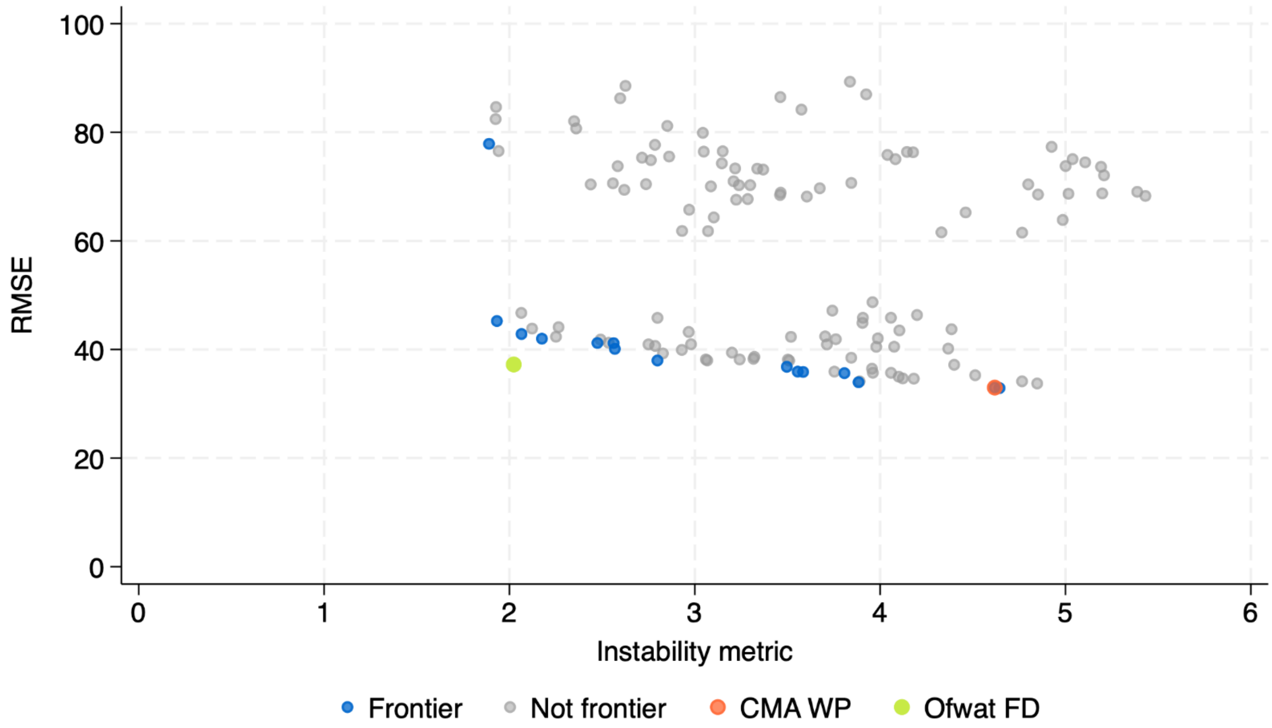
221. We made one restriction to Figure 17, primarily for visual purposes: we excluded models that did not include a scale variable. These models were far worse in terms of fit and stability and their exclusion does not influence the frontier models.

222. We have raised fundamental concerns that the Working Paper's approach to LASSO implementation does not guard effectively against over-fitting, and we see this starkly in Figure 17. We see that the Working Paper's selected model for wholesale water provides slightly better fit than the FD24 model suite but is far more unstable. The Working Paper model seems extremely unstable compared not only to the FD24 models but also to many other possible models that perform similarly in terms of goodness of fit.

223. We present a corresponding chart for wastewater network plus in Figure 18. That shows that the Working Paper's selected model for wastewater network plus is not quite as extreme in terms of instability as for wholesale water. As with its top-down model for water wholesale, the Working Paper's wastewater network plus model has better RMSE than the FD24 wastewater model suite (9% lower) but it has far worse instability (around 90% higher on the instability metric than the FD24 model suite). Furthermore, the Working Paper model is not a frontier model in the sense defined above. For example, we find an alternative model for wastewater network plus which has very similar RMSE to the Working Paper model but which is substantially more stable.

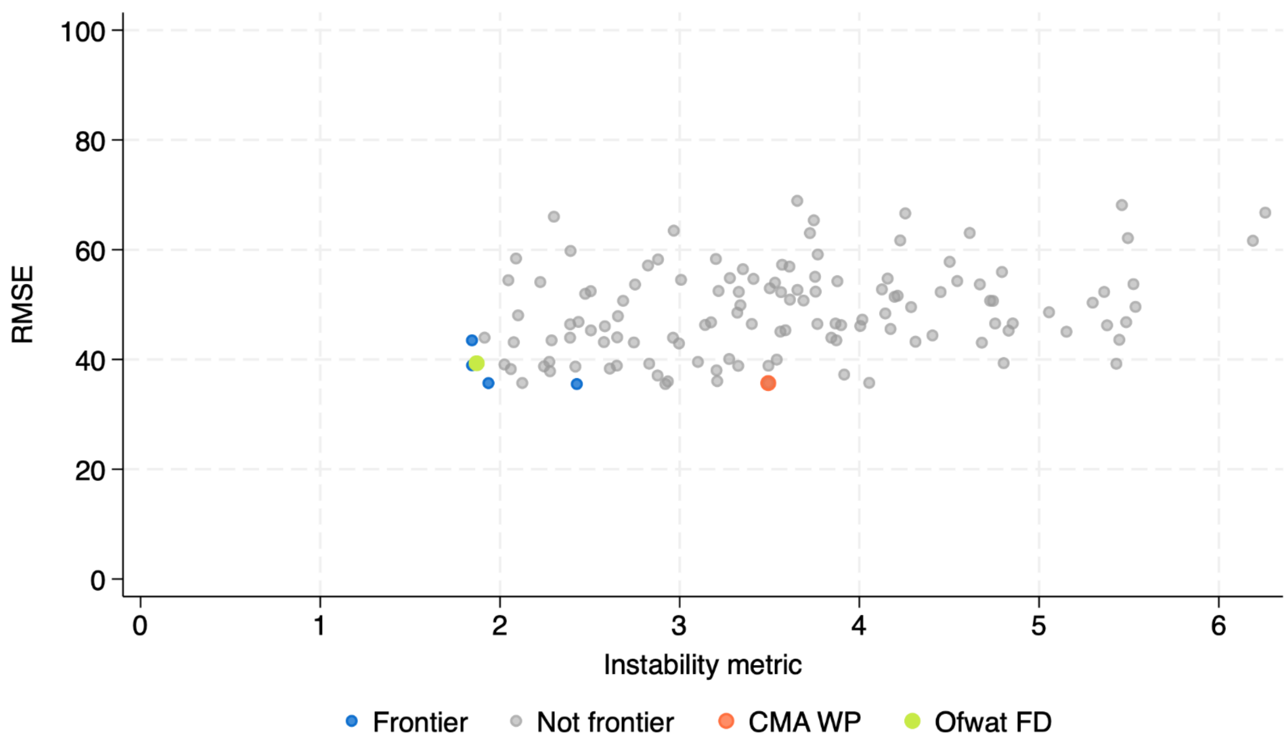
[113] For the FD24 model suite, the results reflect the RMSE and instability of the suite of all the FD24 wholesale water (top-down and bottom-up) models taken together.

FIGURE 17: COMPARISON OF RMSE VS INSTABILITY FOR BROAD SET OF WHOLESALE WATER MODELS



Source: NWL analysis. NWL WPR Databook.

FIGURE 18: COMPARISON OF RMSE VS INSTABILITY FOR BROAD SET OF WASTEWATER MODELS



Source: NWL analysis. NWL WPR Databook.

224. As for the water models, the evidence from this assessment of wastewater network plus models is that the LASSO process has failed to guard against over-fitting and it has failed to select a model that provides a good balance between stability and goodness of fit. While the Working Paper has sought to assert that its model results are sufficiently stable and robust, the analysis is severely compromised by the CMA ignoring the question of whether other models – such as the FD24 model suite or other feasible models – would be more stable and more robust and, in turn, provide a better balance of stability and goodness of fit.

225. The pool of models considered above involve different combinations of explanatory variables from those used by the Working Paper (i.e. initial pool of variables from which the LASSO process selects variables). For consistency with the Working Paper the focus is on OLS models, rather than random effects models, other than for the FD24 model suite which we reproduce models as used by Ofwat which do involve random effects estimation.

6.2.2 The case for reversion to the FD24 model suite

226. We see from Figure 17 and Figure 18 that, compared to a very wide range of possible models, the FD24 model suite strikes a good balance between goodness of fit and stability. This is likely to reflect the favourable attributes of the FD24 *individual* models: models which have been refined and selected using engagement with the industry and insight gained over many years as opposed to models that jump out of an experimental algorithm-based approach. It also reflects the benefits of triangulation across a large and diverse set of models.

227. Our view is that the evidence above (combined with the further testing of the FD24 model suite earlier in the PR24 process) provides a valid evidential basis for the CMA to place 100% weight on the FD24 model suite for its final determinations.

228. We recognise that the CMA had concerns about Ofwat's modelling approach in terms of the treatment of energy prices. We discuss that issue separately in Section 6.3. But it is notable that the FD24 suite of econometric models performs so well in terms of the RMSE of the econometric modelling despite Ofwat's approach which means that changes over time in energy prices are adjusted for after the econometric modelling. In some ways, the charts above *understate* the RMSE performance of Ofwat's overall base cost modelling as the spreadsheet model used for the AMP8 adjustments for energy prices would, if extended back to apply to the historical data period, *reduce* RMSE after energy price adjustments.

229. We also recognise that the CMA using the FD24 model suite for its final determinations would not provide a direct response to each of the individual modelling issues that have been raised by other Disputing Companies. However, the CMA made a choice earlier in its process to focus on a more holistic approach to model selection rather than potential targeted changes to the FD24 models and there is now limited time available to consider the latter at a granular level. The FD24 model suite remains reasonable despite the issues raised by other parties, especially when the merits of

triangulation are recognised. In any event, as we can see from the charts, the FD24 model suite performs well compared to a wide range of top-down models, including the Working Paper LASSO models that had opportunity to select/omit the various variables disputed by other Disputing Companies. If the FD24 model suite had been severely compromised by the issues raised by other Disputing Companies we would not expect the results that we see above.

230. A decision to revert to the FD24 model suite for the CMA's final determinations would be informed by the CMA's exploration of LASSO-based models as a cross-check on the approach and the modelled allowances and the conclusion that they are superior on the overall balance of evidence. This would be consistent with the CMA's objective to "explore a data-driven approach to variable selection using econometric tools such as LASSO" as a means to engage with specific claims put forward by some of the Disputing Companies. The CMA was clear that the use of such a data-driven approach to revise base cost allowances would depend on the "the robustness of the results and the extent to which they differ from Ofwat's".[114] As such, using this analysis to inform a decision to retain the FD24 models is, and always has been, a legitimate outcome of this exercise.

231. Furthermore, the CMA could recommend further investigation of LASSO and/or cross validation techniques in good time for the next water company price reviews, where the ongoing reform agenda already provides a platform for changes to the current approach to cost assessment. We would support the view that water company econometric benchmarking can be improved in the future by making greater use of modern data science techniques, including resampling, and that it would be worth exploring further how algorithms can be used at least for discovery of candidate models even if they are not found to be reliable for the final selection of models.

6.2.3 The potential for triangulation between the Working Paper and FD24 models

232. If the CMA insists on using its LASSO-derived model from its Working Paper for each of wholesale water and wastewater network plus, then at the very least it will need to find a way to triangulate between this model and other (more robust) models. It would not be rational for the CMA, in setting base cost allowances for the Disputing Companies, to place 100% weight on a single LASSO-derived model.

233. In simplest terms, the CMA might triangulate between the FD24 model suite and the Working Paper models. This might be viewed as an approach which combines a model with relatively high goodness of fit with a set of models that are more stable, and which combines a more judgement-based approach to model selection with a more algorithm-based approach.

234. However, we would be highly concerned about the CMA giving, say, 50% weight to the Working

[114] CMA Consultation on Approach and Prioritisation, 28 May 2025, para. 43.

Paper models and LASSO-based approach in any triangulation exercise. There is simply not the evidence to support such a weight on these models. Unless the weight to LASSO models is low (e.g. 10% or 20%) the fundamental concerns we have highlighted about over-fitting and treatment of energy prices would be insufficient resolved. Overall, this approach to triangulation does not seem sufficient.

6.2.4 The potential for triangulation across a different set of models

235. There is also the potential for the CMA to use a triangulation across a set of models that extends beyond the Working Paper models and the FD24 model suite.

236. We have considered various ways that the CMA might identify alternative models to use as part of a triangulation approach. We discuss two options in more detail below:

- a data-driven approach that considers a fuller set of candidate models (see Section 6.2.4.1); and
- models that exclude variables with no non-significant coefficients (see Section 6.2.4.2).

6.2.4.1 Evidence from a data-driven approach that considers a fuller set of candidate models

237. One logical approach we see is to draw on the high-level analysis of goodness of fit and model stability above and select a subset of the “frontier models” to use to set allowances, alongside both the FD24 model suite and potentially the Working Paper models. Of the large number of candidate models we have assessed using objective metrics (RMSE and the instability metric) these frontier models are those for which there is no other model with better performance in terms of both RMSE and the instability metric.

238. If the CMA is determined to retain its LASSO-derived models in some form, these could be used in the triangulation alongside the wider set of models even if they are not frontier models.

239. We considered whether, further to the RMSE and instability metrics, Ofwat’s full set of statistical and diagnostic tests could be used to inform on a selection of a subset of models for inclusion in the triangulation. There is potential merit in that approach if much greater time were to be available, but this would be a complex exercise given the number of candidate modes and there are practical limitations to what can be done given this late stage in the CMA process.

240. Instead, insofar as the CMA chooses to depart from the FD24 models and approach, we suggest that the focus is on the frontier models from the charts above subject to the following restrictions:

- excluding models for which one or more coefficients on explanatory variables have the wrong sign;
- excluding models that have worse RMSE than the FD24 model suite (even if these are more stable). The FD24 models generally have good fit in relative terms, and we can see that a primary motivation of the Working Paper modelling has been to improve fit rather than

stability; and

- excluding models that have worse stability than the Working Paper models: we consider the instability of the Working Paper models is a major concern and the added value of the frontier models is that these could offer a better balance of fit and stability rather than worse stability and even lower RMSE.

241. This exercise is capable of leading to models being used for the triangulation which do not perform as well as the FD24 models under Ofwat's model selection criteria and robustness tests. Our view is that this approach to triangulation would be **sub-optimal compared to the CMA reverting to the FD24 models**. Nonetheless, if the CMA does decide to move away from FD24 model suite, approach and tests (e.g. by including models for which coefficients are not statistically significant at the 10% threshold as the CMA clearly has proposed to do in its Working Paper), it would be **less bad to triangulate across a set of multiple data-driven frontier models as indicated above than to rely on the Working Paper's single LASSO-derived model in isolation.**

242. We set out in Figure 19 for wholesale water, the frontier models that would be selected on the criteria above and compare them to the FD24 model suite in terms of RMSE and the instability metric. We present in Figure 19 the same comparison for the Working Paper model. We then propose weights that might be applied for the CMA's final determination, which would: (a) give weight to both Ofwat's more established model selection approach and more novel and more data-driven approaches; and (b) within the latter, enable triangulation across a set of models that offer improved RMSE versus the FD24 model suite albeit at the cost of some increased instability. We set out the corresponding table for wastewater network plus in Figure 20.

243. Expanding the set of selected models in this way would fit with the CMA's stated aims for model selection to be more data-driven, without placing excessive weight on the unsound and unstable working paper models. An approach along the lines above would allow the CMA to move away from full reliance on the FD24 model suite (despite our view that this should be retained) while allowing for meaningful triangulation across a more diverse set of relatively well-performing models.

FIGURE 19: POTENTIAL BASIS FOR TRIANGULATION: WATER

Model selection approach	Model / model suite	RMSE vs FD24	Instability metric vs FD24	Model weight in WP	Proposed weight for CMA FD*
Ofwat base cost model development process	FD24 model suite	-	-	0%	50%
CMA new LASSO approach	WP model	11% better	127% worse	100%	8.33%
NWL data-driven analysis to identify frontier models in terms of goodness of fit and stability with targeted filtering out	NWL_WW_F1	9% better	91% worse	0%	8.33%
	NWL_WW_F2	4% better	76% worse	0%	8.33%
	NWL_WW_F3	1% better	72% worse	0%	8.33%
	NWL_WW_F4	4% better	87% worse	0%	8.33%
	NWL_WW_F5	4% better	75% worse	0%	8.33%

* Proposal under scenario where CMA has decided not to revert to FD24 models. Our recommended option is for the CMA to revert to FD24 models.
Source: NWL analysis. NWL WPR Databook.

FIGURE 20: POTENTIAL BASIS FOR TRIANGULATION: WASTEWATER

Model selection approach	Model / model suite	RMSE vs FD24	Instability metric vs FD24	Model weight in WP	Proposed weight for CMA FD*
Ofwat base cost model development process	FD24 model suite	-	-	0%	50%
CMA new LASSO approach	WP model	9% better	87% worse	100%	12.5%
NWL data-driven analysis to identify frontier models in terms of goodness of fit and stability with targeted filtering out	NWL_WWNP_F1	9% better	4% worse	0%	12.5%
	NWL_WWNP_F2	1% better	1% better	0%	12.5%
	NWL_WWNP_F3	10% better	30% worse	0%	12.5%

* Proposal under scenario where CMA has decided not to revert to FD24 models. Our recommended option is for the CMA to revert to FD24 models.
Source: NWL analysis. NWL WPR Databook.

244. In contrast, our view is that it would not be rational in light of the evidence presented above (and elsewhere in our response) for the CMA to persist with an approach that gives 100% weight to the LASSO-derived models and 0% weight to either the FD24 models or the frontier models that we have identified in the course of our review of the Working Paper models.

245. We provide further information on the frontier models in Appendix 2 including the coefficients and the results of the various statistical tests in comparison to the WP models.

246. Under the approach to triangulation outlined above, there would be some variation across the models as to whether energy price variables are included in the set of explanatory variables. This is desirable as we do not think that 100% weight should be given to models that involve an energy price variable (as in the Working Paper models). But we would be concerned that some weight would be given to such models. In addition, this approach to triangulation would not amount to a full response to our concerns about over-fitting and instability, though it would be considerably better than using the LASSO models alone.

247. For models that do not include the energy price variable, there would be a need for a post-modelling adjustment as used by Ofwat when setting allowances (see further below for how this could be refined). Similarly, there would be variation as to whether the wage variable is included and, if not, a need for an ex ante adjustment for labour RPEs when setting allowances.

248. We note that, for the frontier water models, there is a mix of models with and without the average pumping head variable. While this was not a selection criterion but it makes sense not to be giving 100% weight to models that do include the average pumping head variable, in light of the material data concerns with that variable that we have set out in Section 5.3.

249. The analysis above is for top-down models and the focus on top-down models reflects the models selected in the Working Paper. In principle, it the same analytical approach could be extended to cover analysis of a large number of feasible bottom-up models based on the set of candidate explanatory variables under consideration. This could inform an assessment of how potential combinations of bottom-up and/or top-down models perform in terms of both goodness of fit and

stability. However, this would be a considerably more complicated exercise. Our focus for the purposes of the response has been on practical improvements to the Working Paper approach, which has selected top-down models.

250. We also considered triangulation between the WP 'top down' and 'bottom-up' LASSO derived models. However, this does not seem a sensible approach as it does not address any of the key problems. The models would still suffer from overfitting and instability, the concerns about modelling energy prices would still not be addressed and with just two models the triangulation would see limited improvement.

6.2.4.2 Models that exclude variables with non-significant coefficients

251. As an alternative to the approach above, a simpler way to try to tackle concerns about over-fitting within the context of the Working Paper models would be to take the LASSO-generated models and drop explanatory variables that are not statistically significant at the 5% or 10% level.

252. The models identified from this exercise could provide a basis for triangulation between the raw LASSO models and versions of the LASSO models that are adjusted to remove variables that are not statistically significant (and for energy make an adjustment outside the modelling). This would be an improvement on the Working Paper approach.

253. However, we would be concerned about a lack of triangulation across alternative models (with different sets of variables) under this approach. Furthermore, given that Ofwat's model selection process has already tried to balance statistical significance against goodness of fit, including through triangulation across a range of more models with different explanatory variables that are statistically significant, we would consider reversion to the FD24 model suite to be superior and to make more sense.

6.2.5 Summary of options relating to triangulation

254. We have identified a range of options for model selection and triangulation in the course of the subsections above. Figure 21 summarises our comparative assessment of how these perform in addressing the key concerns set out at the start of Section 6.

FIGURE 21: COMPARISON OF OPTIONS FOR TRIANGULATION AND HOW THESE TACKLE CORE CONCERNS

Model selection and triangulation option	Addressing concern about reliance on a single model	Addressing concern about over-fitting / instability of results	Addressing concern about CMA energy price modelling
Revert to FD24 model suite in light of analysis of wide range of feasible models	Broad triangulation	FD24 models strike a good balance of fit and stability	Energy prices adjustments applied outside the econometric models
Triangulation across FD24 models, LASSO models and "frontier" models	Broad triangulation (though need to avoid excessive weight to LASSO models)	Better balance of fit and stability than WP, but not as good as FD24	Triangulation would need to give sufficient weight to models that do not include the energy price variable

Triangulation across the set of "frontier" models from analysis of wide range of feasible models	Better than WP, but not as much diversity in models as under FD24	Better balance of fit and stability than WP, but not as good as FD24	Triangulation would need to give sufficient weight to models that do not include the energy price variable
LASSO model subject to removal of explanatory variables that are not statistically significant	Not addressed	Potentially, but may be at the expense of goodness of fit versus FD24 models	Addressed if energy price variables are dropped so that energy adjustments made outside the models
LASSO model with energy variable dropped and post-modelling energy adjustment	Not addressed	Not addressed	Energy prices adjustments applied outside the econometric models
Triangulation between LASSO-derived top-down and bottom-up models	Too few models & not enough diversity across them	Not addressed	Not addressed
Retain WP models (or similar LASSO-derived modes)	Not addressed	Not addressed	Not addressed

Source: NWL Analysis. NB this repeats the information presented in Figure 1.

255. As explained in Section 6.2.2 there are firm grounds for the CMA to use the FD24 model suite for its final determinations, reflecting both the further analysis in our response and the CMA's original statements in May 2025 on its plan to "explore" the LASSO technique.

256. We appreciate that the CMA has concerns with the FD24 model suite. The CMA has, in particular, referred to the Working Paper models performing better in terms of RMSE. But it would be wrong to select a model/approach in a way that focuses on goodness of fit and ignores the relative performance of different models/approaches in terms of the stability of results and/or statistical significance. Furthermore, to the extent that the CMA considers the FD24 model suite to be inadequate in terms of RMSE, the set of frontier models we have identified above provide improvements in RMSE without such an increase in the instability of results as for the Working Paper models.

257. In addition, the CMA has raised simplicity as an objective of its modelling.[115] This cannot possibly be a justification for the approach set out in the Working Paper. The very large amounts of money at stake in the base cost modelling mean it would be wrong to give weight to simplicity (especially if this means using just a small number of models or delegating model selection to an algorithm). Furthermore, the simplest approach would be to focus on Ofwat's established models and approach and consider targeted changes to this. A move to a novel LASSO-based approach, and the introduction of Principal Component Analysis, is the exact opposite of a simple approach.

258. We also recognise that, while we have major concerns about the inclusion of energy variables in the econometric models, based on evidence in this case, the CMA has expressed concerns about

[115] Working Paper, paras 1.5 and 1.6.

Ofwat's approach to energy adjustments. We discuss these issues in the next section.

6.3 POTENTIAL REFINEMENTS TO OFWAT'S ENERGY PRICE ADJUSTMENT

259. We are concerned that the CMA may be reluctant to use the FD24 model suite, despite it performing better overall in statistical terms than the Working Paper models, because of the specific way that Ofwat sought to control for energy prices, which was through a post-modelling adjustment. Furthermore, we are concerned that the CMA might reject models that do not include energy price variables even if these perform relatively well in terms of goodness of fit, stability and the sign and magnitude of coefficients. However, as set out in Section 5.2, the CMA's approach to energy prices leads to unreliable econometric modelling of the relationship between energy prices and costs, which is especially problematic give the scale of energy price changes over the sample period.

260. In principle, an alternative to the approach involving explanatory variables for energy prices is an approach involving pre-modelling adjustments to expenditure input data to take account of energy prices. This provides an alternative way to control for energy price changes than relying on estimated coefficients from including energy prices as an explanatory variable, while tackling some of the CMA's apparent concerns with Ofwat's approach (e.g. impacts of energy price impacts on the profile of efficiency scores over time). But this is a complex exercise and there would be a need for both pre-modelling adjustments to the expenditure data and for post-modelling adjustments to predicted costs to take account of the energy price forecasts for AMP8. There is considerable scope for implementation error. This might be something that the CMA could recommend for consideration for future price reviews when a longer timeframe is available to develop and test the approach.

261. Ofwat's approach of making purely post modelling adjustments for energy prices has the benefit of relative simplicity while also avoiding the risks arising under the Working Paper's approach.

262. While the CMA has expressed concerns in its Working Paper about the profile of efficiency scores under the FD24 models, the practical implications of this issue are in danger of being overstated. If efficiency scores really are treated as an accurate guide to efficiency/inefficiency, it does not make sense for these to increase over time. But that would be a misplaced view of what efficiency scores represent: despite the terminology, they capture a wider range of factors (see Section 3.1). If efficiency scores are found to increase over time, this should not automatically be interpreted as revealing increases in inefficiency; it is more likely to point to factors that have affected costs over time that are outside of the econometric models. Our view is that these factors include, for example, increases in costs associated with service quality improvements over time (i.e. companies responding as intended to ODI incentives) which are not captured in the modelling. Leaving aside the treatment of energy prices, we would expect efficiency scores to increase over time given the factors (such as service quality) omitted from the models.

263. This point aside, there could be two refinements to Ofwat's approach, and there are interactions between this issue and other parts of our case to consider.

264. First, as set out in our SoC we consider that under Ofwat's FD24 there was an inconsistency between: (i) making adjustments for an upper quartile efficiency score calculated over five years from 2019/20 to 2023/24; and (ii) making an adjustment for energy prices by comparing assumed AMP8 energy prices to the average energy prices over the whole period.[116] This issue is closely related to our submission that what base buys for water mains renewal should be calculated by reference to average renewal rates in the period from 2019/20 to 2023/24 rather than over the full sample period.[117] This issue also applies if, as we propose, the CMA moves away from an adjustment based on upper quartile efficiency score to one based on the median efficiency score.

265. Assuming that the CMA has accepted our case on what base buys then, for internal consistency, it would make sense to use the last five years of energy price data in the calculation of the ex post adjustment for energy prices. For our SoC, we had calculated revisions to the CMA's energy price adjustment to use the five-year period from 2019/20 to 2023/24 to calculate the average electricity price reflected in (post-UQ) allowances. This resulted in reductions to our allowances of £42m against the Ofwat FD24 allowances, as reported in our SoC.[118]

266. Second, given the more detailed review of energy price coefficients in Section 5.2 above, we think that it would make more sense to deduct an estimate of the energy cost share of CPIH from the power cost share of each company, when calculating the adjustment. While there is considerable uncertainty about that we refer to a figure of 6.6% from the ONS in Section 5.2.

[116] NWL SoC, para 347.
[117] The reasoning for that is set out more fully in NWL PD Response Section 3.1 on what base buys for mains renewal which covers a wider set of considerations.
[118] NWL SoC, para 349.

7 RPES AND COST ADJUSTMENT CLAIMS

267. Section 5 of the Working Paper concerns true-ups for energy and labour/wages and the approach to certain cost adjustment claims (relating to wages and energy as well as economies of scale at water treatment works and coastal population). We discuss the following issues in turn:

- RPEs (see Section 7.1); and
- cost adjustment claims on economies of scale and coastal population (see Section 7.2).

268. The Working Paper does not cover cost adjustment claims related to asset health. As such we have not made further comments on this matter here, which is addressed in detail in our PD Response.[119]

7.1 RPES

269. The Working Paper provides a minor update on the PD's approach to true ups for energy and wages. Whilst this provides some minor clarification about the formula, further details are required to understand how it will operate and there are still many unaddressed issues with the PD that need to be acted upon.

270. In particular, for labour RPEs:

- **ex ante labour RPEs for modelled base costs**: through the inclusion of the wage variable in the top-down water model and the forecast of this variable in AMP8, the Working Paper does provide ex ante RPEs for water but not for wastewater. As per our PD Response, an ex ante adjustment is required even when wages are not part of the model to ensure we can recover our efficient costs.[120] For wastewater, we could have to wait seven years until the true up mechanism is operated before we can recover efficient costs in this area. The overall, approach is such a significant departure from well-established regulatory precedent that it needs to be corrected;
- **forecasts used for construction wages**: the Working Paper still bases these on the CMA's own forecasts. These are less appropriate than OBR wage forecasts which take better account of future economic impacts on wages;[121]
- **regional and sector mix of wages**: the Working Paper still assumes all wages should go in line with regional construction wages. As per our PD Response, we do not see how this matches the reality of operating a water company where there is a mix of skills from different sectors and some jobs are not location dependent and are therefore better proxied by national indices;[122] and

---

[119] NWL PD Response, Section 3.2.
[120] NWL PD Response, Section 2.3.1.
[121] NWL PD Response, Section 2.3.3.
[122] NWL PD Response, Section 2.3.5.

- **labour RPEs for other areas of base costs**: whilst not the focus of the Working Paper, RPEs for unmodelled costs and network reinforcement need to be included in the CMA's final determination.[123]

271. While we welcome the further detail on the true up which is a core report of the regulatory framework, we still think it requires further thinking to make it operational and consistent with the remainder of the approach to base costs:

- **the "% contribution" for energy in the true up formula**: the Working Paper is not clear about how the power cost share should be determined. For internal consistency with the approach taken for efficiency scores (whether used for an upper quartile or median adjustment) we believe that the power cost share for the true-up should be determined as a single industry-average figure using the reported power cost shares for the five-year period from April 2019 to March 2024. This is consistent with the approach used by Ofwat in FD24;
- **what the "change in index" means**: the description of the true-up in the Working Paper suggests that the adjustment would be for "changes in the relevant index […] over time".[124] The CMA should clarify what it means by this. We believe that adjustment should account for differences between forecast and out-turn values of the relevant index over AMP8, rather than changes over time in that index; and
- **application of the wages true up to wastewater**: It is not clear from the Working Paper whether the CMA intends for the wages true-up to apply to both water and wastewater, or just for water. We strongly recommend that the CMA provides up front labour RPE allowances for both water and wastewater and that the ex post true-up adjustment to applies to both water and wastewater. As is the case with the energy, it is not clear how the labour cost share should be determined. For consistency, we recommend that the CMA uses the labour cost shares that Ofwat had intended to use for its labour RPE mechanism as set out in FD24.[125]

7.2 COST ADJUSTMENTS CLAIMS ON ECONOMIES OF SCALE AND COASTAL POPULATION

272. In relation to water treatment works economies of scale, the Working Paper suggests adding the two extra explanatory variables relating to WTW economies of scale to the pool of variables fed into the LASSO process for the bottom-up WRP model.[126] Our analysis indicates that this would lead to all eight explanatory variables being selected for the WRP model. This would exacerbate our concerns about over-fitting and unreliable modelling. This does not seem a valid basis for setting or adjusting allowances.

---

123 NWL PD Response, Section 2.3.2.
124 Working Paper, para 5.3.
125 Ofwat PR24 Final Determinations Expenditure Allowances, Table 33.
126 Working Paper, paras 5.8 and 5.12.

273. Similarly, adding the coastal population to the bottom-up STW model would lead to all seven candidate explanatory variables being selected by the LASSO process and would lead a model for which the wage variable has the wrong sign. As above, this would exacerbate our concerns about over-fitting and unreliable modelling. This does not seem a valid basis for setting or adjusting allowances.

274. Across both areas, we are also concerned about the potential inconsistency of the CMA using bottom-up models to calculate cost adjustments applied to allowances set by the top-down models. There is the potential for this to lead to perverse effects.

275. In relation to coastal population, the CMA has asked whether it would be possible to construct a robust coastal load time series. In the time available for this response, we have not formed a firm view on this issue, but our more general concern is that the CMA's focus should be on addressing the fundamental deficiencies of its modelling approach rather than looking to create new variables at this stage of the process.

8 APPENDIX 1: FAILURE TO PROVIDE AN EXPLANATION FOR THE REDUCTION IN BASE COST ALLOWANCES UNDER THE CMA'S MODELS

276. The Working Paper proposes a novel approach to model selection, presents models that suffer from quality concerns, and then opts to rely on just a single model for each of water and wastewater. This results in modelled costs that are materially lower than Ofwat's FD24 allowance, whichever comparison is used:

- the Working Paper's AMP8 modelled costs (post UQ adjustment) are **-5.0%** for water and **-6.8%** for wastewater compared to FD24;[127] and
- the change in total modelled costs for the whole sector are -**7.2%** for water and **-7.0%** for wastewater.[128]

277. The IWC has been very clear in expressing its concerns about the implications of over-reliance on a "data-driven, econometric approach" for ensuring an appropriate level of cost allowance for each individual company.[129] In that context, and noting the real world challenges with the plausibility of delivering such material reductions, it is essential that CMA understands what is driving the lower cost allowances under its models and can adequately explain that difference and why it is appropriate.

278. According to the Working Paper, the primary source of difference at the industry level is the Working Paper's more stretching upper quartile efficiency adjustment, driven by energy costs:

> The more stretching catch-up challenge produced by our model compared to Ofwat's model is largely due to the inclusion in our model of energy and other input prices. Specifically, through this mechanism, our model can account for the effect of the sustained shock to energy and other input prices from 2020/21 onwards that contributed to water companies overspending their PR19 allowance.[130]

279. As set out in the following sub-sections, this does not explain the size of the difference in the allowances. As such, the Working Paper risks misleading the CMA panel (and other stakeholders in this process) about the reasons why the Working Paper's experimental LASSO modelling approach has led to lower allowances across the industry than Ofwat's FD24. Instead, the Working Paper has simply explained why the upper quartile efficiency scores are systematically higher, across the industry, under the FD24 approach than the Working Paper. The CMA has not properly engaged with, or understood, why the Working Paper models result in substantially lower allowances than the FD24 models. This undermines the credibility of its results and reduces the

[127] NWL analysis based on Working Paper. Average across industry of % changes between CMA updated allowances and Ofwat FD allowances.
[128] Working Paper, Tables 3.4 and 4.3. The slightly larger reduction for the industry aggregate compared to the average across companies reflects larger reductions for some of the larger companies.
[129] Independent Water Commission final report, Independent Water Commission Final Report, para 417. See also NWL PD Response, Section 2.2.
[130] Working Paper, para. 3.21.

reliance that should be placed upon them.

8.1 OFWAT'S APPROACH TO THE CALCULATION OF THE UQ EFFICIENCY SCORES

280. Ofwat uses the terminology of an upper quartile catch-up efficiency challenge (or adjustment). But in practice, Ofwat applies an adjustment to the predicted values from the econometric models which takes account of both: (a) differences in costs between companies which are not explained by its models; and (b) differences in costs over time that are not explained by its models. Under Ofwat's approach:

- **efficiency scores are calculated over last five years:** Ofwat calculates "efficiency scores" using data for the five-year period to 2023/24 (i.e. the last five years of data used). It takes each company's actual costs over that five-year period divided by the modelled costs predicted by models estimated on the full historical dataset (before applying any post-modelling adjustments to modelled costs); and
- **selection of upper quartile benchmark:** The adjustment is based on the upper quartile of the efficiency scores that have been calculated over the last five years.

281. The first element above has the effect of indirectly adjusting for changes in costs over time that are not captured by the econometric modelling: if companies' costs have increased over time compared to the modelled predictions, efficiency scores will tend to increase over the period. This is what we see in Figures 3.2 and 4.2 in the Working Paper: the average efficiency score under Ofwat's modelling is well above 1 for each of the last five years.

282. Seen in this context, the terminology of efficiency scores is potentially misleading. An increase in efficiency scores does not imply worsening efficiency over time, as it may be driven by other factors not captured in Ofwat's models (e.g. input prices or increased costs to deliver performance improvements – see Section 3.1). These factors mean that differences in efficiency scores between companies, or over time, are not a reliable guide to differences in efficiency between companies or over time.

283. Ofwat's efficiency scores are calculated *before* taking account of its post-modelling adjustment for energy prices. Ofwat made separate post-modelling adjustments for energy prices during AMP8. Had it applied its energy price adjustment approach to modelled costs over the historical dataset, rather than just AMP8 modelled costs, it could have calculated efficiency scores after the adjustments for energy prices. That would have taken better account of changes over time in energy prices and the profile of average efficiency scores would have been more stable over time (though not entirely flat as other factors not captured in the models, besides energy prices, may be pushing up costs).

8.2 THE CMA FAILS TO COMPARE EFFICIENCY SCORES AND ALLOWANCES ON A LIKE-FOR-LIKE BASIS

284. The fact that the Working Paper models include the electricity price variable, while the FD24 models do not, helps explain why the efficiency scores are higher under than the FD24 models than under the Working Paper. The CMA has in turn provided a narrow technical explanation of why the upper quartile efficiency adjustment is greater under the Working Paper models compared to the FD24 models.[131] But this comparison of the Working Paper and FD24 efficiency scores is misleading because it is not on a like-for-like comparison. The calculation of the upper quartile efficiency score under the Working Paper models is made *after* taking account of energy prices whereas under Ofwat's approach it is made *before* taking account of energy prices (a separate step that Ofwat applied after the upper quartile adjustment).

285. The difference in approach to energy prices does not explain the difference in allowances. Both the approach of including energy prices as an explanatory variable and an approach based on making adjustments outside of the modelling are methods to take account of changes in energy prices over time when setting allowances for AMP8. There is no a priori reason why differences in the approach to energy prices should cause such big reductions to allowances (e.g. 6-7%) across the industry.

8.3 DIFFERENCES IN EFFICIENCY SCORES BETWEEN COMPANIES DO NOT EXPLAIN THE SIZE OF THE REDUCTION

286. While the CMA seeks to explain the reduction in allowances by reference to the "stronger" efficiency challenge under the Working Paper models,[132] the differences in efficiency scores between companies do not explain the extent of the difference. For instance, we can compare the upper quartile to median efficiency scores under the Working Paper and the FD24 models. As shown in Figure 22 below, the Working Paper models involve a greater difference between median and upper quartile efficiency scores compared to the FD24 models. This can explain some of the reduction in industry-wide allowances but the scale of differences (e.g. of the order of 1% or 2% of base costs) does not come close to explaining the extent of reduction in sector-wide cost allowances under the Working Paper approach.

FIGURE 22: DIFFERENCE BETWEEN UPPER QUARTILE AND MEDIAN EFFICIENCY SCORES

	Water		Wastewater	
	Working Paper	FD24	Working Paper	FD24
UQ efficiency score	0.93	0.99	0.94	0.99
Median efficiency score	1.01	1.04	0.97	1.01
Difference: UQ minus median	0.08	0.06	0.03	0.02

Source: Working Paper; FD24, NWL Analysis

[131] Working Paper, paras. 3.21-3.24.
[132] Working Paper, para 1.9.

9 APPENDIX 2: FRONTIER MODELS

287. In Section 6.2.4 we present a range of frontier models that provide a better trade-off between RMSE and stability compared to the CMA WP models. These models are presented in Figure 19 and Figure 20 as the basis for potential triangulation.

288. Figure 23 and Figure 24 below present the variables, coefficients and statistical tests for these models. The results show that in addition to the improved trade-off between RMSE and stability shown in Section 6.2.4 the models generally perform stronger than the Working Paper's models against the standard measures of statistical significance, and sensitivity. However, the FD24 model suite is still better overall than this set of alternatives.

FIGURE 23: FRONTIER MODELS – WATER

Cost driver	Explanatory variable	NWL_WW_F1	NWL_WW_F2	NWL_WW_F3	NWL_WW_F4	NWL_WW_F5	WP WW
Scale	Scale combined (log)	-0.678***	-0.683***	-0.678***	-0.680***	-0.692***	-0.683***
Complexity	Weighted average treatment complexity (log)	0.516***	0.541***	0.555***	0.375*		0.450***
Topography	Average pumping head TWD (log)				0.155	0.232	0.096
	Booster pumping stations per length of mains (log)	0.325**	0.368***	0.335**			0.318**
Density	Density combined (log)	0.057***	0.046**	0.059***	0.027	0.031	0.050***
	Density combined (log) - squared	0.025***	0.024***	0.025***	0.032***	0.035***	0.026***
Input price	Energy index (log)	0.153*			0.171**	0.206***	0.116*
	Construction wages (log)		0.487				0.303
Constant	Constant	4.818***	4.337***	5.535***	2.893***	2.969***	3.804***
Robustness tests	Adjusted R-squared	0.971	0.97	0.969	0.968	0.965	0.971
	RESET test	0.036	0.004	0.016	0.108	0.55	0.065
	VIF	1.968	2.327	1.962	1.316	1.124	2.4
	Pooling	0.999	0.852	0.902	1	0.999	0.994
	Normality	0.291	0.16	0.096	0.018	0.023	0.139
	Heteroskedasticity	0	0	0	0	0	0
Sensitivity tests	Removal of the most efficient company	Green	Green	Green	Amber	Amber	Amber
	Removal of the least efficient company	Green	Amber	Green	Green	Green	Amber
	Removal of first year	Green	Green	Green	Green	Green	Green
	Removal of last year	Amber	Green	Green	Green	Green	Amber

Source: NWL Analysis. NWL WPR Databook.

FIGURE 24: FRONTIER MODELS - WASTEWATER NETWORK PLUS

Cost driver	Explanatory variable	NWL_WWNP_F1	NWL_WWNP_F2	NWL_WWNP_F3	WP WWNP
Scale	Scale combined (log)	-0.301***	-0.305***	-0.303***	-0.307***
Economies of scale	Weighted average treatment size (log)			-0.028	-0.018
	Load treated in size bands 1 to 3 (%)				0.004
Complexity	Load treated with ammonia consent <=3mg/l (%)	0.006***	0.005***	0.006***	0.006***
Topography	Pumping capacity per sewer length (log)	0.480***	0.513***	0.453***	0.467***
Urban rainfall	Urban rainfall per sewer length (log)	0.110**		0.116**	0.115**
Input price	Energy index (log)	0.169	0.193*	0.167	0.167
Constant	Constant	4.844***	4.390***	5.140***	5.013***
Robustness tests	Adjusted R-squared	0.949	0.943	0.95	0.95
	RESET test	0.102	0.79	0.169	0.194
	VIF	4.107	4.04	4.724	5.071
	Pooling	0.991	0.994	0.987	0.999
	Normality	0.053	0.009	0.083	0.065
	Heteroskedasticity	0.023	0.052	0.026	0.033
Sensitivity tests	Removal of the most efficient company	Amber	Green	Green	Red
	Removal of the least efficient company	Green	Green	Red	Red
	Removal of first year	Green	Green	Green	Green
	Removal of last year	Amber	Amber	Amber	Amber

Source: NWL Analysis. NWL WPR Databook.