Water PR24 References

The CMA base cost modelling working paper – UUW consultation response

January 2026

This document contains UUW's comments in response to the CMA's Base cost modelling working paper in relation to the Water PR24 References made by Ofwat on behalf of disputing companies.

United Utilities

Water for the North West

Executive Summary

This document sets out United Utilities Water (UUW)'s representations in relation to the CMA's Base costs modelling working paper for the PR24 Water References.

The CMA redetermination is an important process that supports the accountability and effectiveness of the economic regulatory framework. UUW is committed to ensuring that the regulatory regime works well for customers, investors and the environment. To this end, we are providing representations to the CMA's working paper to ensure that any decision that might inform future regulatory periods supports this objective. We would like to thank the CMA for the opportunity to respond to its proposal and for its transparency in providing additional coding detail used for its econometric modelling.

Although UUW has not asked Ofwat to refer its final determination to the CMA, the approach that the CMA takes to setting its determinations, particularly to assessing efficient base costs, can have a wider long-term consequence for the whole industry, including UUW.

We have significant concerns about the proposed use of machine learning techniques, such as LASSO (Least Absolute Shrinkage and Selection Operator), for deriving base allowances. This is a novel and untested approach in the context of setting cost allowances for utility sectors. Whilst we are not opposed in principle to the development of new and improved estimation techniques, upon detailed review of the CMA's base models, we note the proposals represent a very significant **U-turn against previous regulatory cost assessment methodologies**, which have centred around appropriate economic and engineering principles rather than purely statistical driven approaches. **The change of approach also runs counter to the general principle of regulatory consistency and predictability, and if it were to be adopted at future price controls, we consider it would likely lead to suboptimal outcomes for customers and the environment by misallocating the resources required by companies to deliver their functions in delivering a vital public service.**

We have specific concerns regarding the implied reduction in base allowances across the water industry relative to Ofwat's PR24 Final Determinations. The implied reduction would be around £3bn, at a time when the Independent Water Commission (IWC) has called for a greater focus of resources on asset health and resilience. This is a further stretch of approximately £1bn relative to the provisional determination published in October. We find this movement in allowances incongruent with the findings of both the Independent Water Commission (IWC)'s report - which has recommended a greater focus on asset health and resilience – as well as Ofwat's own work in implementing a new Asset Health cost change item to allow companies to deliver greater investment in capital maintenance.

Whilst we appreciate that the CMA should not be constrained by the IWC's findings for the purpose of the CMA's PR24 price determinations, its approach to developing these base models and the weight placed on them, to set appropriate allowances is in stark contrast to both the IWC's conclusions and the wider sectoral consensus. Across the industry, there is a growing recognition that the activities funded through base expenditure require enhanced support if companies are to maintain and improve the resilience of their networks and assets.

Our review of the models has focused on cross referencing the CMA's methodology against Ofwat's cost assessment methodology and UUW's principles of regulatory cost assessment, as set out in previous published papers to develop the approaches used at PR24. The approach adopted is therefore a long standing consistent and principles-based approach.

In particular, the new base models present four fundamental issues:

(1) **A lack of interpretability and transparency**, with the individual contributions of some cost drivers being obscured by the revised approach adopted by the CMA in the working paper. Companies and stakeholders will struggle to verify whether their specific cost pressures are being reflected in allowances. Relationships between cost drivers have been implicitly introduced without any economic or engineering justification.

(2) **An over-reliance on a single model, without triangulation across a diverse suite of models**, such as the 32 models used in Ofwat's FD. The CMA's revised approach fails to sufficiently reflect the heterogeneous operational characteristics faced by companies, leading it to incorrectly attribute additional expenditure incurred as 'inefficiency'. Therefore using one single 'best' model for cost assessment is unlikely to deliver a robust and fair allocation of allowances for each individual company, ultimately to the detriment of customers. We note that the importance of reflecting company-specific factors was explicitly highlighted by the IWC as an area for increased regulatory consideration in future.[1] The CMA's revised approach, by contrast, reduces the scope for recognising legitimate differences in operating environments, and therefore is moving further away from this recommendation.

(3) **The appropriateness of LASSO in the context of setting base allowances in the water industry.** LASSO adds further complexity to cost assessment without delivering meaningful variable selection, as the CMA has restricted the candidate variable set to a narrow suite of pre-determined cost drivers—some of which are further compressed into composite variables. In this low-dimensional setting, LASSO simply selects every available variable, offering limited insight into the true drivers of cost for each service, consequently reducing transparency and interpretability.

(4) **A data-led modelling approach undermines regulatory certainty and long-term stability of base allowances.** Relying heavily on LASSO introduces volatility into model specifications and, in turn, base allowances—particularly in a sector with small datasets and limited variation. This reduces predictability for companies and investors, which is contrary to the IWC's call for a simpler and more stable regulatory framework. Model instability risks an increase in the cost of capital, constrains long-term asset planning, and risks poorer outcomes for customers and the environment.

We appreciate that the CMA has sought to address some of the concerns listed in our initial response, in particular by:

- **Removing the interactions between wage, energy and scale in its revised econometric models;**
- **Performing statistical robustness tests** to its revised econometric models to ensure that the models perform well statistically; and
- **Rectifying multicollinearity concerns** through the introduction of composite variables combining multiple cost drivers into a single variable**.**

However, given that the main issues identified in our response remain unresolved, we ask that the CMA:

- **Recognises the limitations of a purely data driven approach** to developing base models. This would likely involve considering whether **machine learning models, such as LASSO, should be used more as a diagnostic and investigative tool with appropriate caution**, assisted by expert judgment and economic intuition. LASSO and similar tools could be used alongside (rather than as an alternative to) the established cost modelling approaches to identify new cost drivers that might have not been detected through existing models; and
- **Considers utilising a more diverse suite of models** to better reflect the differing operational circumstances faced by companies, which cannot be fairly captured within one (or at most two) single model(s).

We would also recommend that the CMA undertakes an external peer review of these models to test for accuracy and consistency of the models given the untested and novel approach adopted in the provisional determinations.

Whilst we welcome the development of new improved estimation techniques, following our detailed review of the new base models developed by the CMA since its previous PDs, **we believe there are fundamental issues with the new models that mean they would be unsuitable for use at future price reviews for the purpose of setting base costs allowances.** As a result, we consider that the CMA should proceed with considerable caution if it is minded to apply them in these redeterminations.

[1] https://assets.publishing.service.gov.uk/media/687dfcc4312ee8a5f0806be6/Independent_Water_Commission_-_Final_Report_-_21_July.pdf, page 193, para 417

Contents

Appendices

1. Introduction

1.1 Background and context

United Utilities Water (UUW) is the appointed water and wastewater provider for the North West of England. Our purpose is to provide great water for a stronger, greener and healthier North West. We serve more than three million homes and 200,000 business across a region that stretches from Carlisle in north Cumbria to Crewe in south Cheshire, whilst safeguarding 1,300 km of coastline and 7,000 km of rivers.

The North West is a diverse area with more than seven million people living across both major metropolitan areas such as Greater Manchester and Merseyside as well as small rural villages in the heart of Cumbria. To deliver our services to customers we manage hundreds of reservoirs, treatment works and pumping stations and more than 120,000 kilometres of water pipes and sewers.

Like the appellant companies in this case, UUW is regulated by Ofwat – amongst other regulators – and is subject to Ofwat's price review determinations.

1.2 Why we are responding to the CMA's base cost working paper

The CMA determination is an important process supporting the transparency, accountability and effectiveness of the regulatory framework. We are committed to ensuring that the regulatory regime works well for customers, investors and the environment and, to this end, we are providing representations to the CMA's base costs working paper to ensure that any decision that might influence future regulatory periods supports this objective.

Although UUW has not asked Ofwat to refer its final determination to the CMA, the approach that the CMA takes to setting its determinations, particularly to assessing efficient costs, can have wider long-term consequences for the whole industry, including UUW. To this end, UUW's representations are aimed at ensuring that any potential change to future cost assessments improves the accuracy, appropriateness and fairness of the allocation of funding allowing each individual company to meet their obligations to the benefit of customers and local communities.

UUW submitted an initial response to the CMA's provisional determination. Whilst many of the issues raised in that document remain unresolved, heightened concerns regarding interpretability and the further reduction in LASSO's suitability for variable selection warrant this updated response paper. The issues we raise in this paper are generally consistent with, and build upon, the points we raised in our response to the provisional determination. Where the CMA has updated its approach, our response takes this into account.

We do not seek to opine on the merits or otherwise of the specific circumstances that individual companies may represent to the CMA or have represented previously to Ofwat. For the most part, our observations in this submission focus on the potential impact that these decisions may have on future price control and reflect long-held policy positions and evidence promoted by UUW through Ofwat's engagement with the industry in the development of the PR24 methodology.

We set out our key representations in Section 2.

2. Representations

2.1 Base cost models

There are four key issues associated with the new base cost models developed by the CMA.

2.1.1 A lack of interpretability and transparency

**The CMA's working paper exacerbates the interpretability concerns we raised in in our previous response to the provisional determination[2]. ** The introduction of composite variables (see section A.2.2 for our detailed discussion of composite variables) obscures the individual contribution of underlying cost drivers, making it difficult for companies and stakeholders to assess whether the econometric models accurately reflect the cost economic and engineering cost relationships that companies experience.

This lack of interpretability also makes it difficult to test whether the variables are intuitive and consistent with economic and engineering rationale. In addition, the use of squared composite variables creates implicit interactions between certain cost drivers for which there is no clear economic or engineering rationale.

A further concern is that the loadings used to construct the composite variables have been derived using a linear assessment, despite the expectation of a quadratic relationship. This inconsistency makes it highly likely that the loading values are incorrect, but the lack of interpretability makes it very difficult for companies to assess the impact of this.

2.1.2 An over-reliance on a single model rather than a diverse suite of models

We suggest that the CMA reconsiders what we believe to be an over-reliance on a single set of models (one for wholesale wastewater and one for wholesale water) rather than adopting a suite of models with different variable specifications, as Ofwat did at PR24. This is particularly important given the inherent limitations of any econometric modelling technique when applied to small samples with limited variability in the explanatory factors.

While LASSO statistical models have been widely used to develop forecasts of the average dependent variable in large datasets (for example the average forecast of stock prices or house prices), predictive accuracy is not the goal in a regulatory price control. A LASSO type approach is likely to explain efficiency differences *within the model*, rather than predicting underlying efficient costs. Given the small dataset available in the water sector, the likelihood of spurious correlations is significantly higher than in larger datasets, increasing the risk that the model explains efficiency differences within the model in order to predict costs. This is not adequately aligned to the actual need in a price control to predict efficient costs for each company, not to provide the best prediction of actual costs for each company.

The CMA has itself recognised the issues caused by small datasets. The CMA has previously stated that "*any model must be seen as an approximation*"[3] and *"… all econometric models are imperfect, and it is not possible to establish with certainty that they incorporate every single determinant of costs."*[4]. Although the CMA's single-model approach may improve predictive accuracy relative to historic data, it does so only for an average company. It therefore does not address or improve the ability to address the issue of companies' heterogeneity and operational characteristics such as topography and geology. By contrast, using multiple models with different variable specifications allows for a more accurate reflection of individual companies' operational environments, as demonstrated in Appendix A.

[2] The CMA Provisional Determinations – UUW consultation response.
https://assets.publishing.service.gov.uk/media/6941883a1ec67214e98f302e/United_Utilities_17.12.25.pdf

[3] Competition and Markets Authority (October 2015).
https://assets.publishing.service.gov.uk/media/56279924ed915d194b000001/Bristol_Water_plc_final_determination.pdf, pg. 59: para. 4.9 and pg.73: para. 4.50.

[4] Competition and Markets Authority (2025).
https://assets.publishing.service.gov.uk/media/6836bd365f8330ed48e72b24/PR24_Approach_and_Prioritisation.pdf, pg. 13: para. 46.

Furthermore, the use of a single model removes fundamental cost drivers. For example, although density is a well-established driver of sewage collection costs, it has not been selected in the CMA's top-down wastewater model. This omission risks producing unduly stringent allowances and introduces significant omitted-variable bias.

Water companies have vastly complex and varying configurations of assets and operations, necessary to meet the quality standards and service level expectations across the different operating environments across England and Wales. Given the small number of companies, and hence the small dataset to inform any econometric assessment of costs, it would be optimistic to expect that one single model could be capable of adequately and fairly assessing efficient costs for all companies. No single model specification can fully capture the cost drivers necessary for each company to meet its quality and service level obligations.

We recommend that, in line with the PR24 approach adopted by Ofwat, **the CMA adopt a suite of models that use different types of variables to capture key cost drivers, with results triangulated across the different models**. This approach provides a fairer and more robust basis for setting efficient allowances, as it better reflects the operational characteristics of individual companies. Triangulation across multiple models also offers a more transparent and contestable way of accommodating differences between companies.

If the CMA chooses to proceed with its LASSO-based approach, it should do so with caution, given the issues identified to date in its modelling.

2.1.3 Appropriateness of application of LASSO

The CMA's application of LASSO is problematic for two reasons: it is unsuitable for assessing company-specific cost adjustment claims, and it cannot function effectively within the constrained modelling framework adopted by the CMA in this revised approach.

Firstly, cost adjustment claims arise precisely where industry-wide models fail to capture unique company-specific operational circumstances. LASSO, however, is designed to identify variables with broad explanatory power across the sector on average. It will naturally overlook drivers that are only relevant for a small number of companies, even when these are legitimate and material. This reinforces the need for post-modelling adjustments to ensure that engineering-led, company-specific cost pressures are appropriately recognised. Assessing such claims solely through a model optimised for the "average" company risks systematically under-allowing efficient costs for certain companies.

Second, LASSO works best when assessing which variables, from a large and diverse pool, best explain 'average' movements. However, cost assessment is focused on predicting efficient costs at a company level. The 'average' cannot be thought to accurately predict each individual company's circumstances. It is entirely possible that the variable selection would be different under an approach which prioritises 'average' model fit versus one which is focused on company appropriateness. As such, we consider LASSO only adds complexity to the modelling methodology, without improving the modelling outcome.

2.1.4 A data-led modelling approach undermines regulatory certainty and long-term stability of base allowances

The CMA's entirely data-driven approach to model development reduces regulatory certainty around future base allowances. This is reflected in the instability of company-specific allowances observed in sensitivity testing.

Reduced predictability runs counter to the principles of stability set out in the government's Better Regulation Framework[5] and, ultimately, limits companies' ability to plan and deliver sustainable asset maintenance so critical for a long-term sector like the water industry. The predictability of the regulatory regime is also a key consideration for credit rating agencies, increasing the cost of capital and, ultimately, customer bills. This means that, if the CMA chooses to continue using the LASSO approach mechanistically to select the explanatory variables

[5] As set out in Department for Business and Trade (September 2023). https://assets.publishing.service.gov.uk/media/67587ba55a2e4d4b993bfa83/better-regulation-framework-guidance-2023.pdf.

for base models in its final determinations, it is unlikely to be suitable for setting price controls at future periodic reviews.

By design, LASSO technique produces different selections and combinations of cost drivers, coefficient signs, and allowances over time, depending on which variables marginally improve the statistical fit to historic data. This effect is further exacerbated by the industry's inherent small dataset and limited variance in explanatory variables over time. As a result, using LASSO to develop base models generates greater instability than Ofwat's PR24 base models, as demonstrated in Appendix A.

Considering the empirical investigations presented in Appendix A, we consider that the current base models lack the necessary robustness to provide confidence in their capability to set appropriate base allowances at the next price review. The limited interpretability of the model coefficients suggests that the selected variables—and the magnitude of their effects—may not accurately reflect the underlying cost drivers in the sector.

It is important to recognise that the ultimate objective of any robust cost assessment should not necessarily be model optimisation and statistical fit, but rather a fair and accurate allocation of funding to ensure each individual company is able to meet its obligations. However, the new base models have been developed primarily through a statistical process aimed at maximising fit on a purely data-driven basis, without coherent testing against the well-established economic and engineering relationships observed in the industry in practice.

Absent such testing, while the **new models might perform better on a purely average statistical predictive basis, they may reflect spurious relationships that are not grounded in the underlying cost structure of the water industry**. **A good data fit does not guarantee alignment with efficient sector costs, especially given factors such as inefficiency, varying costs across periods, and maintenance cycles. This risks inappropriate variable selection and suboptimal outcomes for allocation of capital, to the detriment of customers, the environment and both the performance and investability of the sector.**

The CMA's model suite also appears to be in stark contrast with previous determinations by the CMA where Ofwat's models were criticised for their shortfalls and their incapability "*to interpret the relationships that they imply between costs and explanatory variables in economic and engineering terms*[6]".

We provide a detailed discussion of each of these issues and supporting evidence in Appendix A.

[6] Competition and Markets Authority (October 2015). https://assets.publishing.service.gov.uk/media/56279924ed915d194b000001/Bristol_Water_plc_final_determination.pdf, pg. 72: para. 4.50.

Appendix A

The CMA's approach to assessing base costs is entirely novel within regulatory cost assessment in the English and Welsh water sector. It uses a machine learning technique called LASSO to identify an appropriate set of cost drivers for use within the botex model suite. In doing so, it overrides Ofwat's model suite which, whilst not perfect, is the product of almost a decade of cross-industry collaboration and development. In particular, the focus on backward-looking statistical model fit undermines a focus on engineering, economic and operational rationale within the model suite.

We acknowledge the CMA states in its working paper that '*coefficients are not the primary goal of the modelling exercise and may not always lend themselves to unique economic interpretation*'. However, the CMA's proposed approach seems ill suited for the main purpose of regulatory cost assessment. Cost assessment should not seek the best performing statistical model. Instead, the end goal is an allocation of cost that is broadly appropriate for each company's individual circumstances in a future regulatory period (or AMP). **It is entirely possible that models that achieves this does do not 'fit' historic data better than an alternative – but we are ultimately interested in the ability of the model to make predictions of future costs. As such, we are sceptical of an approach which uses LASSO in isolation from industry knowledge and experience.**

This appendix gives an overview of the analysis we have carried out in response to the CMA's proposed approach to assessing base expenditure allowances.

A.1.1 Overview of our approach

We requested the code that the CMA used to implement its revised approach, which allowed us to examine its methodology in detail. We would like to thank the CMA for demonstrating a commitment to regulatory transparency through its provision of this code. We also recognise and appreciate the CMA's efforts to resolve some of the initial concerns raised, from the implementation of statistical tests to attempts to mitigate multicollinearity concerns, as well as removing the interaction terms from wage and energy.

Our approach has focused on examining the CMA's methodology against Ofwat's cost assessment methodology and UUW's principles of regulatory cost assessment, as set out by UUW in a methodology paper to inform and develop the approaches used at PR24[7]. The approach adopted here is a long standing and principled approach and aligns to the approach in our response to the CMA's PD. In particular, this has involved:

- Considering how intuitive and interpretable the CMA's methodology and model suite is.
- The appropriateness of LASSO in this context.
- Considering how coherent the CMA's framework is.
- Testing the robustness of the CMA's models and examining how this compares to Ofwat's models' robustness.

The following sections outline our assessment of the CMA models against a set of principles developed over the last two regulatory periods by both Ofwat and water companies. These principles have not been developed to prescribe a single best technical solution to cost assessment, but to provide an objective framework grounded in sound economic and engineering rationale to assess methodological choices transparently, recognising that no one-size fits all approach can appropriately reflect the complexity of the sector.

We consider it appropriate to apply these principles given the CMA's adoption of LASSO – a novel approach to cost assessment in the context of the water industry and, as far as we know, untested in the UK regulated utility sectors for the purpose of assessing efficient costs and setting appropriate allowances. Its use by the CMA marks a significant departure from previous methodologies and introduces a fundamentally different way of modelling base expenditure allowances.

This assessment allows us to evaluate whether the CMA's methodology aligns with an objective cost assessment framework, which was designed to cope with changes in technical approaches to cost assessment. Importantly,

[7] United Utilities (July 2021). https://www.unitedutilities.com/globalassets/documents/pdf/the-principles-of-regulatory-cost-assessment.pdf.

the principles are not intended to identify the 'best model' in a purely statistical sense, but to support the assessment of future efficient costs for each company.

While LASSO is objective in that it is determined by algorithms, it focuses solely on maximising statistical fit in the historic dataset. We consider that it is better to define an objective framework that can best achieve a suitable base cost allowance for each company in future periods. A purely backwards-looking approach (such as LASSO) cannot, in our view, achieve this.

A.1.2 Ofwat's cost assessment framework

Firstly, we have reviewed the base models developed by the CMA and have applied the following tests consistently with the framework developed by Ofwat and that underpin a robust cost assessment modelling approach:

- **Stability and robustness:** Are the estimated model results stable / robust to changes in the underlying assumptions and data (e.g. different sample period; alternative model specification)?
- **Intuitive and interpretable:** Are the estimated coefficients of the right sign and of plausible magnitude?
- **Predictive power:** Can the models suitably assess efficient costs?
- **Statistical validity:** How do the models perform across a range of statistical diagnostic tests?

Table 1 below outlines our assessment across the wholesale water and wastewater The following sections investigate each dimension in more detail.

Table 1: Assessment of CMA models against Ofwat's principles

Model	Stability and robustness	Intuitive and interpretable	Predictive power	Statistical validity	Pass/fail
Water					**Fail**
Wastewater					**Fail**

Source: UUW analysis of CMA publication

As set out in table 3.5 and 4.5 of the CMA's working paper[8], the revised models developed by the CMA show statistical test results that appear to be more robust than under CMA's provisional approach. As such we have now assessed statistical validity as green for both models.

A.2 A lack of interpretability and transparency

The new base models exhibit counterintuitive features; the coefficients of several variables lack interpretability.

A.2.1 LASSO results in unusual model specifications, increasing complexity and reducing transparency and interpretability

Ofwat's base models incorporated a range of different cost drivers within each model, typically using a single variable to proxy for each cost driver. It varied the choice of proxy across its suite of models and then triangulated the results. This approach helped maintain simplicity within individual models and enabled stakeholders to assess transparently whether the estimated coefficients were of a reasonable sign and magnitude.

Our prior concerns have only been amplified by this new approach – with model the suite now reduced from three models to just two. This further limits the variation in operational characteristics that can be captured through the models, rather than expanding it, as we previously argued was necessary.

[8] CMA (2025) Base Costs Modelling Working Paper. Available at https://assets.publishing.service.gov.uk/media/6943c024143d960161547e21/Base_Costs_Modelling_Working_Paper..pdf. Section A.17.

The table below summarises the key issues identified at both the Provisional Determination stage, and the working paper stage, together with an indication of whether each issue has been resolved.

Issue	PD	Working paper	Comment
Small number of top-down models	✓	✓	Not resolved, model suite made smaller than PD
Size of variable set	✓	✓	Number of candidate variables has decreased further. LASSO failing in its purpose of cost driver selection.
Complexity of composite variables		✓	Composite variables introduced to resolve issues of multicollinearity but raise concerns around interpretability.
Interaction terms	✓	✓	Whilst wage and scale interaction terms have been removed, the squaring of principle components introduces new, unjustified, interactions.
Coefficient magnitude	✓		Coefficients no longer directly interpretable due to the use of principle components.
Coefficients of wrong signage	✓		Resolved where interpretable.

A.2.2 Lack of interpretability of composite variables

Transparency is a key pillar of effective regulation and is especially important in the context of the CMA's proposed framework in its revised working paper. The move towards the use of composite variables significantly reduces stakeholders' ability to understand how base cost allowances are derived. Composite variables, by design, combine multiple underlying cost drivers into a single variable. Whilst this may address some concerns regarding multicollinearity – *the scenario where explanatory variables are highly correlated, making it difficult to isolate their individual effects* - it obscures the individual contribution of each driver and makes it far more difficult for companies, or other interested stakeholders, to understand how operational characteristics are being reflected in the model.

This loss of visibility directly affects companies' ability to understand the reasons why they may appear efficient or inefficient in the model. When several cost drivers are merged into a single variable, this approach does not show the contribution of each underlying cost drivers preventing stakeholders from assessing whether this composite measure is intuitive and accurately reflecting the underlying economic and engineering cost relationships. This lack of visibility is particularly problematic in a sector with such a diverse range of operating conditions.

This contrasts with Ofwat's approach to base modelling, where each cost driver is represented explicitly, allowing stakeholders to assess whether the sign, magnitude and rationale for each variable are reasonable. This has

helped to facilitate the ongoing dialogue between companies and regulators in this area over a sustained period of time.

The original variable coefficients can, in principle, be derived, however doing so requires extracting the loadings and standard deviations from the underlying code. The CMA has not provided the required method for transforming the squared principal components, which prevents full reconstruction of the coefficients.

A.2.2.1 The existence of interaction terms

Whilst the linear PC1 variables can be back transformed once the relevant loadings and standard deviations have been extracted from the code, the nonlinear PC1 variable, notably density in the wholesale water model, requires a far more complicated transformation, and implicitly assumes that interactions exist between the original variables.

In practice, interaction terms are introduced, yet the CMA does not acknowledge this in its response. The inclusion of $PC1^2$ does not only account for the u-shaped relationship with density, but implicitly embeds interactions between the variables. For example, by using this variable, there is an assumption that the effect of weighted average density depends on properties per length, and that the strength of this interaction is structurally constrained by the linear loadings. See the equation below for the mathematical illustration of this point.

Figure 1: Depiction of a non-linear principal component [9]

Definition of the first principal component:

$$PC_1 = \sum_{i=1}^{3} \alpha_i \tilde{X}_i$$

Definition of the squared first principal component:

$$PC_1{}^2 = \left(\sum_{i=1}^{3} \alpha_i \tilde{X}_i\right)^2 = (\alpha_1 \tilde{X}_1 + \alpha_2 \tilde{X}_2 + \alpha_3 \tilde{X}_3)^2 = \alpha_1{}^2 \tilde{X}_1{}^2 + \alpha_2{}^2 \tilde{X}_2{}^2 + \alpha_3{}^2 \tilde{X}_3{}^2 + 2\alpha_1\alpha_2 \tilde{X}_1 \tilde{X}_2 + 2\alpha_1\alpha_3 \tilde{X}_1 \tilde{X}_3 + 2\alpha_2\alpha_3 \tilde{X}_2 \tilde{X}_3$$

The CMA has not provided any economic or engineering rationale for why an interaction should exist between the variable *lnWAD_MSOAtoLAD_population* and *lnpropperlength* within the wholesale water density principal component. This lack of explanation is important, because the PCA loadings imply a substantive relationship between these two variables. Using the coefficients published by the CMA, the implied interaction is as follows $2\alpha_1\alpha_2 = 2 \times 0.579 \times 0.572 = 0.331$.

However, there is no discussion of why weighted average density and properties-per-length should interact in this way, nor why such a relationship would be expected based on underlying engineering characteristics or cost drivers.

The presence of this interaction also undermines the interpretability of the coefficients, as the existence of interaction terms means that a variable's marginal effect will be conditional on the other variable. To evaluate these effects consistently against the data, a specific value—such as the mean—of the interacting variable must be selected, adding an additional layer of complexity.

Furthermore, the loadings for each variable in PCA are derived under the assumption that the variables enter the analysis in a linear form. When a principal component is squared after PCA, these original loadings no longer represent meaningful weights because the transformation introduces nonlinearities and interaction terms among the original variables. Consequently, the interpretation of the loadings is invalidated. The loadings of $\alpha_i{}^2$ and $\alpha_i\alpha_j$ were never optimised by PCA. Indeed, there is a risk that this method overloads on one variable which is the best fit on a linear basis and then forced this into squared variable, without considering whether this variable would be

[9] CMA (2025) Base Costs Modelling Working Paper. Available at https://assets.publishing.service.gov.uk/media/6943c024143d960161547e21/Base_Costs_Modelling_Working_Paper..pdf. Section A.17.

weighted as such once in squared form. This appears to be a clear inconsistency in approach. However, given the lack of interpretability, we are unable to fully assess the impact of this error on the model results.

A more appropriate approach would have been to include both the linear and nonlinear transformations of the original variables prior to constructing the principal components. This ensures that the PCA loadings reflect contributions from both linear and nonlinear forms of the variables, rather than imposing a nonlinear structure after the fact.

Finally, the existence of these interaction terms makes it particularly difficult to derive the coefficients of the original variables from the squared principal component. The CMA has not provided an equation by which to do this, it is difficult therefore for UUW as a third party to derive this.

A.2.2.2 The exclusion of the other principal components

Moreover, the CMA has just included PC1 in the variable set presented to LASSO. Whilst this approach is more parsimonious, it is a further example of a strong data-fitting approach. PCA determines components based solely on maximising the variance of the independent variables, without considering their relationship to the dependent variable. There is no guarantee that the components with the highest variance (PC1, PC2, etc.) or their squared terms will be the most predictive of the outcome.

As stated by the CMA, on page 2, paragraph A.12 of its working paper, '*Subsequent principal components capture the next largest amount of remaining variance, but with a crucial constraint: it must be completely uncorrelated (orthogonal) with the first component. The third principal component is orthogonal to both the first and second, and so on. This orthogonality means each component captures independent information – there is no redundancy between them, unlike the original variables*'. This implies that there may be a key aspect of the cost drivers influencing costs which is not included using exclusively PC1.

This is a further example of a redundant use of LASSO by restricting the variable set a priori.

A.2.2.3 Cost adjustment claims cannot be assessed in an environment with uninterpretable coefficients

Cost adjustment claims are a critical feature within the regulatory framework, designed to ensure that companies are funded for legitimate cost pressures that cannot be captured within the scope of standard econometric models. By definition, these claims require a clear understanding of how the model treats individual cost drivers, so that companies can demonstrate where their circumstances differ from the assumptions embedded in the model. For this process to work, it fundamentally relies on the ability to interpret model coefficients and understand the relationship between specific cost drivers and allowed expenditure.

The CMA's proposed modelling approach—built around composite variables—removes this interpretability, as discussed in A.2.2.1. In this environment, cost adjustment claims become very difficult to assess. A company cannot demonstrate that a particular driver is not captured by the model if the model does not allow a transparent interpretation of how that driver is treated and therefore is performing within the model.

Equally, the regulator cannot meaningfully test whether a cost adjustment claim is already reflected in the model, because the coefficients do not correspond to interpretable variables. This creates a circular problem: the regulator rejects claims on the basis that the model "does not select" the driver, yet the model itself provides no interpretable evidence to support or refute that conclusion.

The implications extend beyond the cost adjustment process. As an example, a modelling framework built on uninterpretable coefficients also undermines the ability to determine "what base buys", which is a fundamental requirement for Ofwat's recently implemented asset health cost change item which relies on a clear understanding of the level of asset maintenance and renewal activity funded through base allowances, so that any incremental pressures or shortfalls can be assessed correctly. If the underlying models use composite variables whose coefficients cannot be easily interpreted, regulators and companies cannot identify which cost drivers are being funded through base, nor the scale of funding associated with each driver. This makes it challenging to operate a credible and evidence-based asset health reopener.

A.3 An over-reliance on a single model, without triangulation across a diverse suite of models

A.3.1 Lack of model triangulation

All models are imperfect representations of reality. A single imperfect model can be acceptable in certain circumstances (e.g. when attempting to identify the causal impact of a particular variable). In the context of setting base cost allowances for individual companies, the risk of an imperfect model must be taken seriously and appropriately mitigated. This has led us to promote the use of a diverse and triangulated model suite. Ofwat adopted a similar approach in its PR24 model suite.

A core principle of Ofwat's approach to cost assessment is the triangulation of results across a diverse suite of models. This is particularly important in the water sector, where the sample size is inherently small and companies operate in highly varied environments. No single model can fully capture the full range of operational, geographic and asset-related cost drivers across the industry. Triangulation mitigates some of these model limitations by averaging across multiple model specifications, reducing the risk that any one model's structure or variable selection disproportionately influence allowances. A key example of the problematic nature of using a single model to represent Wastewater is the resultant exclusion of any density driver, despite it being a key cost driver for sewage collection.

At PR24 Ofwat used a broad suite of 32 models and triangulated the results to ensure that no individual model unduly benefited or penalised any particular company. This diversity of models helped to provide a more balanced assessment of efficient costs across the sector. In this context, the CMA's revised approach represents a significant step backwards. The CMA has now adopted only two top-down models—one for Water and one for Wastewater—compared with the 32 models used by Ofwat. This reduction in model diversity substantially increases the risk that the chosen model specifications will not adequately reflect the cost relationships faced by all individual companies. Rather than mitigating model uncertainty through triangulation, the CMA's approach amplifies it by placing almost all weight on a very narrow set of model structures.

Importantly, there has actually been a deterioration in this area relative to the Provisional Determination. Rather than strengthening the modelling framework in response to stakeholder concerns, the CMA has reduced the number of models further, increasing reliance on a small number of top-down specifications. This makes the modelling less resilient and less capable of reflecting company specific operating conditions.

A.3.2 One model cannot accurately predict allowances for all companies

A key feature of regulatory cost assessment in the water sector is that companies' regions are not homogenous. In this context, we would not expect that a single model can accurately predict allowances for all companies, especially given the small cross-sectional dataset. This heterogeneity led Ofwat to a diverse model suite that mitigates the risk that a single model unduly benefits or penalises certain companies and minimises the risk of substantial incidence effects.

It also led Ofwat to supplement its modelling with company-specific adjustments for unique circumstances e.g. through cost adjustment claims. By definition, a cost adjustment claim is required when a company faces cost pressures from a factor that cannot be captured by the model. It is therefore not clear why the CMA has chosen an approach where it rejects company claims on the basis that the modelling has deemed that the driver is not important.

In contrast, LASSO has been widely used to develop predictive models (for example for stock or house prices) with large datasets and large pools of potential explanatory variables, where the focus is on the average prediction.

While LASSO can be effective in identifying variables that enhance statistical fit and improve predictions for the efficient costs of the average company, using it mechanistically to construct an industry-wide model will, by design, capture only the cost relationships that hold on average. **As such, it may not be suitable for reflecting the cost relationship that exists at a company level, which is what we consider to be the key purpose of the regulatory cost assessment. We note that the importance of company-specific factors was explicitly highlighted**

by the IWC as an area for increased regulatory consideration in future, compared with the approach Ofwat had historically taken, including at PR24. [10]

A.4 Appropriateness of use of LASSO

A.4.1 Cost adjustment claims are unsuitable to be assessed through LASSO

Cost adjustment claims are by their nature, company-specific, and arise precisely because standardised econometric models are unable to fully capture the unique cost pressures faced by individual companies in certain areas. Attempting to resolve these issues through the use of LASSO is counter-intuitive, given that LASSO is designed to identify variables that best explain cost variation across the industry as a whole.

LASSO prioritises variables with broad explanatory power, potentially overlooking factors that are critical for one, or a subset of companies but not statistically significant at the aggregate level. We would therefore expect that some legitimate cost drivers may not be selected by LASSO, despite being materially relevant to individual companies. This reinforces the importance of retaining mechanisms such as cost adjustment claims, that ensure that engineering-led company-specific circumstances are appropriately recognised and allowed for.

The CMA rejects bespoke cost claims or avoids engaging fully with the underlying rationale for operational specific characteristics. **In our view, not all company specific adjustments can be captured within the scope of the benchmarking model alone.** Post modelling adjustments are used where the model is not thought to be capable of capturing legitimate company or time specific variation in costs. They can serve as a necessary backstop in regulatory cost assessment and provide companies with an avenue to ensure all efficient cost pressures are reflected appropriately.

While we do not comment on the legitimacy or otherwise of the claims presented to the CMA, **we recommend that the CMA recognises the potential shortcomings of an approach in assessing company-specific claims through this benchmarking approach**.

A.4.2 LASSO fails to achieve its purpose

A fundamental requirement for the effective use of LASSO is a large pool of potential explanatory variables from which the algorithm can select. LASSO is specifically designed for high-dimensional settings, where the number of candidate variables is large relative to the sample size, allowing the tool to identify the subset of variables that best explains variation in the dependent variable. This is not the case in the CMA's proposed framework. By restricting the variable list to those historically used by Ofwat, the CMA has not expanded the set of potential cost drivers beyond an already narrow suite, limiting the scope for LASSO to operate as intended.

The dimensionality of the dataset is reduced even further through the use of composite variables. By collapsing multiple scale and density drivers into a small number of principal components, the CMA has constrained the model to a pre-determined set of inputs rather than providing LASSO with a richer and more diverse variable pool. This is the opposite of the environment in which LASSO is designed to function.

The CMA suggests that limiting the variable list to that established by Ofwat in the FD ensures compliance with engineering and economic rationale and maintains consistency with Ofwat's modelling framework. However, this argument highlights the inappropriateness of LASSO in this context. If the variable list is intentionally restricted to a small number of pre-validated cost drivers, then the key advantage of LASSO—its ability to sift through a large number of potential variables and choose the most relevant—is not actually being fully utilised. In effect, the CMA could reasonably be characterised as applying a method designed for high-dimensional modelling to a low-dimensional dataset, and then further reducing that dimensionality through the creation of composite variables.

This raises the fundamental question about the suitability of LASSO for regulatory cost assessment in the water sector. A method that requires a large and diverse variable set to function effectively cannot be expected to perform well when applied to a deliberately constrained set of inputs, in a sector with a particularly small sample

[10] https://assets.publishing.service.gov.uk/media/687dfcc4312ee8a5f0806be6/Independent_Water_Commission_-_Final_Report_-_21_July.pdf, page 193, para 417

size and heterogeneous operating environments. The result is a modelling framework that does not exploit the strengths of LASSO, while simultaneously introducing weaknesses around interpretability and consistency.

The CMA itself defines LASSO as 'An econometric/machine learning technique which selects explanatory variables that best predict the outcome variable of interest.' [11], However, as the CMA has only provided a limited variable set for consideration, LASSO provides no insight into which drivers matter most. Indeed, LASSO has selected all of the candidate variables proposed in both water and wastewater top-down models. As such, LASSO introduces additional complexity without performing its core function.

Table 2 below demonstrates that every variable made a candidate for consideration in the wholesale water model has been selected.

Table 2: LASSO's choice of variables from the candidate set in the wholesale water model.

Candidate Independent Variables	Inclusion in water model
(log) Scale combined (PC1)	✓
(log) Density combined	✓
Squared (log) Density combined (PC1)	✓
Weighted Average Complexity (WAC) (log)	✓
Booster Stations per length of mains (log)	✓
Average Pumping Head (TWD)	✓
Hourly regional construction wages (log)	✓
DESNZ energy price index (log)	✓

Source: Scripts 2A and 3A of CMA provided code, Table 3.1. of working paper.

This is also the case for wastewater, where every candidate variable has been chosen by LASSO. Despite its importance as a cost driver for SWC (sewage collection) the CMA did not include density within its candidate variable set for WWNP, and thus removed LASSO's role of deciding its importance. The potential for a cost driver that has wide industry support (e.g. density) to be excluded —whether due to the CMA's initial candidate set or LASSO's variable selection—highlights a fundamental limitation of relying on a single top-down model rather than adopting a triangulated approach. A triangulated approach allows cost assessment to reflect the impact of different cost drivers, as opposed to taking a binary 'in/out' approach.

The CMA's preference for a model with superior statistical fit, i.e. the top-down wastewater model - creates a material risk that important SWC-specific cost drivers may be overlooked. While the top-down approach may perform well in aggregate, because it is built on sector-wide averages. It smooths over the heterogeneity that exists between companies, particularly those with distinct operating environments.

As a result, factors that are clearly relevant cost drivers for SWC, but do not manifest strongly at the industry-wide level, may be excluded or under-weighted simply because they do not improve the overall statistical performance of the single best model. This reinforces the need for triangulation across modelling approaches: reliance on a single "best-fit" model can obscure genuine cost-causal relationships that are only visible in bottom-up models or company-specific analyses, and therefore risks producing allowances that are misaligned with the true cost pressures companies are actually facing.

[11] CMA (2025) Provisional Determinations Volume 5: Appendices A-F and Glossary. Available at https://assets.publishing.service.gov.uk/media/68e7c17cb6ae95e3c76907c6/PR24_PD_vol_5_-_appendices_and_glossary.pdf. Page 61

Table 3: LASSO's choice of variables from the candidate set in the wastewater model.

Candidate Independent Variables	Inclusion in wastewater model
(log) Scale combined (PC1)	✓
Load treated with ammonia consent <3mg/l	✓
Pumping capacity per sewer length (log)	✓
Urban rainfall per sewer length (log)	✓
Hourly regional construction wages (log)	✓
Weighted average treatment size	✓
Load treated in size bands 1 to 3 (%)	✓
DESNZ energy price index (log)	✓

Source: Scripts 2B and 3B of CMA provided code, Table 4.1. of working paper.

A.3.9 LASSO's selection of cost drivers

A purely statistical approach tends to favour variables with simple, strong empirical relationships, even when these are not the most conceptually appropriate. In some cases, a variable with a weaker statistical fit may have a far stronger economic or engineering rationale. A clear example in water is energy costs. It is straightforward for a model to identify a direct relationship between energy use and costs—for instance, through pumping head. Companies that rely heavily on groundwater abstraction typically incur higher pumping costs and therefore appear more sensitive to the energy index.

However, groundwater is generally of higher raw quality than surface water and requires less complex treatment. The true cost relationship is therefore not linear or simple: high energy use may be offset by lower treatment costs. A model that prioritises the strong statistical relationship with energy consumption, whilst rejecting the weaker but conceptually important driver of treatment complexity, risks systematically favouring high-energy-consuming companies.

This example illustrates how a data-driven approach such as LASSO can skew outcomes by oversimplifying cost relationships. **By selecting variables solely on the basis of predictive strength, the model may embed biases that do not reflect the underlying engineering and operational realities of the sector failing to be effective in setting base allowances that reflect the true scope for efficiency.**

A.3.10 Non-linear relationships

LASSO selects variables based on their ability to reduce prediction error, rather than on whether they represent a genuine nonlinear relationship. The inclusion of a squared density term may therefore function more like a RESET-type diagnostic, signalling that the original specification omitted nonlinearities or interactions.

In practice, squaring a compound variable introduces implicit interactions—for example, between the variable *lnproperties* and the variable *lnlengthsofmain*—meaning that LASSO's selection of this term **may simply be capturing these artificially constructed interactions rather than identifying a true (quadratic) relationship with density.**

A.3.11 The risk of spurious correlation and LASSO

Given the small sample size, any trend in company efficiency that correlates with the scale variable could distort the estimated cost relationships. In such cases, the model becomes primarily data-fitting rather than conceptually driven.

Under LASSO's selection criteria, variables that are correlated with inefficiency may appear to have stronger explanatory power for costs, even if this relationship is spurious. **This risks penalising efficient companies whose underlying cost drivers differ from the dominant pattern in the data.**

A.5 A data-led modelling approach undermines regulatory certainty and long-term stability of base allowances

We do not consider that a data-led approach supports regulatory certainty. It also risks the stability of future allowances and companies' ability to plan activity across regulatory periods.

As recommended by the IWC's recent report *"The regulator needs to make the (economic) framework simpler, less aggressive, more predictable, more realistic and easier for investors to understand and forecast"*[12]. **The predictability of the regulatory regime is a key consideration for credit rating agencies and will, ultimately, be a significant driver of the cost of capital. It is also important to recognise that the water sector is a long-term industry, and companies require certainty and stability around base expenditure allowances to be able to plan and manage the maintenance of long lived and sizeable assets.**

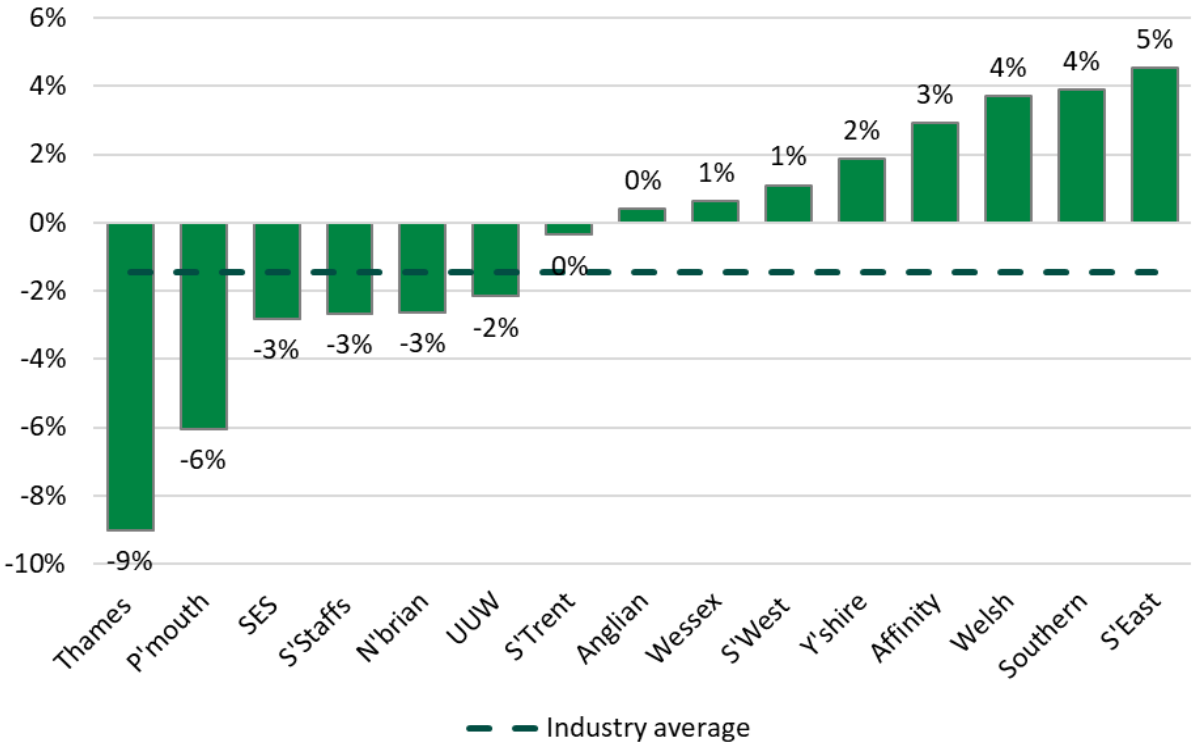
We therefore consider that if the LASSO approach were to be adopted beyond this determination—without further testing against established economic and engineering rationale—it would materially reduce the certainty and predictability of companies' base allowances. This could increase the cost of capital and, in turn, customer bills. Reduced certainty in allowances would also constrain companies' ability to deliver sustainable and cost-effective maintenance programmes, to the detriment of customers and the environment.

Figure 2 below illustrates the relative stability in company totex allowances (pre-catch up) between Ofwat's FD and the CMA's revised approach. Whilst the overall movement in industry allowances is relatively small at -1.5%, the variation in company allowances are far more pronounced, ranging from a -9% decrease and a 5% increase. This degree of volatility highlights the importance of assessing the impact of model changes not only at an aggregate level but also on each company individually. Significant swings in allowances can have material implications for companies' ability to deliver their statutory functions.

This point is particularly relevant in the context of the CMA's use of LASSO-based statistical models. Although LASSO techniques are well-established for forecasting average outcomes in large datasets, these applications rely on extensive amounts of data and are designed to identify a single best-fit model for an "average" case. Regulatory cost assessment has a fundamentally different purpose. The objective is not to optimise predictive accuracy for an average company, but to set efficient allowances that reflect the diverse operating environments faced by each individual water company that is subject to a determination.

[12] Independent Water Commission (July 2025). https://assets.publishing.service.gov.uk/media/687dfcc4312ee8a5f0806be6/Independent_Water_Commission_-_Final_Report_-_21_July.pdf, pg. 326: para. 755.

Figure 2: Movement in totex allowances from Ofwat's FD to revised CMA working paper



A.5.1 CMA dismissal of dropping a company as a valid method of testing model stability

If a merger were to occur and the number of wastewater companies were reduced, the basic regression requirement that N (the number of observations) must exceed k (the number of variables plus the constant) would no longer be met. In practical terms, once a company is removed, there is insufficient independent variation for the regression to function. This highlights a broader structural issue: the modelling framework is overly dependent on statistical techniques as the end-point of allowance setting, despite the sector's inherently small sample sizes and limited cross-sectional and panel variation.

The CMA's approach—particularly its use of the Swamy-Arora estimator—relies critically on the current number of companies and variables being maintained. **Any reduction in the dataset, such as through merger activity, would undermine the feasibility of the model and further expose the fragility of a methodology that depends on such a limited dataset.**

This supports the use of a simple, triangulated model suite similar to the one Ofwat used at PR19 and PR24.

A.5.2 A strictly data-led approach such as LASSO is unstable

LASSO is a technique used to select variables from a defined pool of variables and determine the appropriate specification for each variable (e.g. linear, quadratic, interactive), based on the impact on the Root Mean Square Error (RMSE). Applying LASSO mechanistically to determine base models could lead to different selections and combinations of cost drivers and parameters over time, depending on which variables best explain variance in the data at that point.

LASSO is typically most effective in high-dimensional environments where there are many potential regressors relative to the number of observations. When applied to small samples with only a limited set of regressors, issues of consistency and interpretability arise. A purely data-led approach introduces instability into model predictions: as the data changes, the model specification changes with it, creating uncertainty around long-term allowances. This instability is likely to be exacerbated by the small dataset in the water sector, where minor changes in inputs can lead to disproportionately large variations in results.

Machine-learning techniques such as LASSO may have value as diagnostic or exploratory tools, helping to identify potential new cost drivers that traditional approaches might overlook. However, using LASSO as the primary mechanism for determining regulatory base models risks embedding volatility rather than providing stability.

In a regulatory context, the sensitivity of LASSO to new years of data would lead to frequent model changes from one AMP to the next. This would make it difficult for companies to plan and invest with confidence, given the material shifts in base allowances that could arise purely from changes in model specification. Crucially, we do not consider that the CMA has provided a robust explanation for the significant reduction in industry base allowances implied by its approach, nor how this aligns with the wider cost pressures currently facing the sector.

United Utilities Water Limited
Haweswater House
Lingley Mere Business Park
Lingley Green Avenue
Great Sankey
Warrington
WA5 3LP

unitedutilities.com