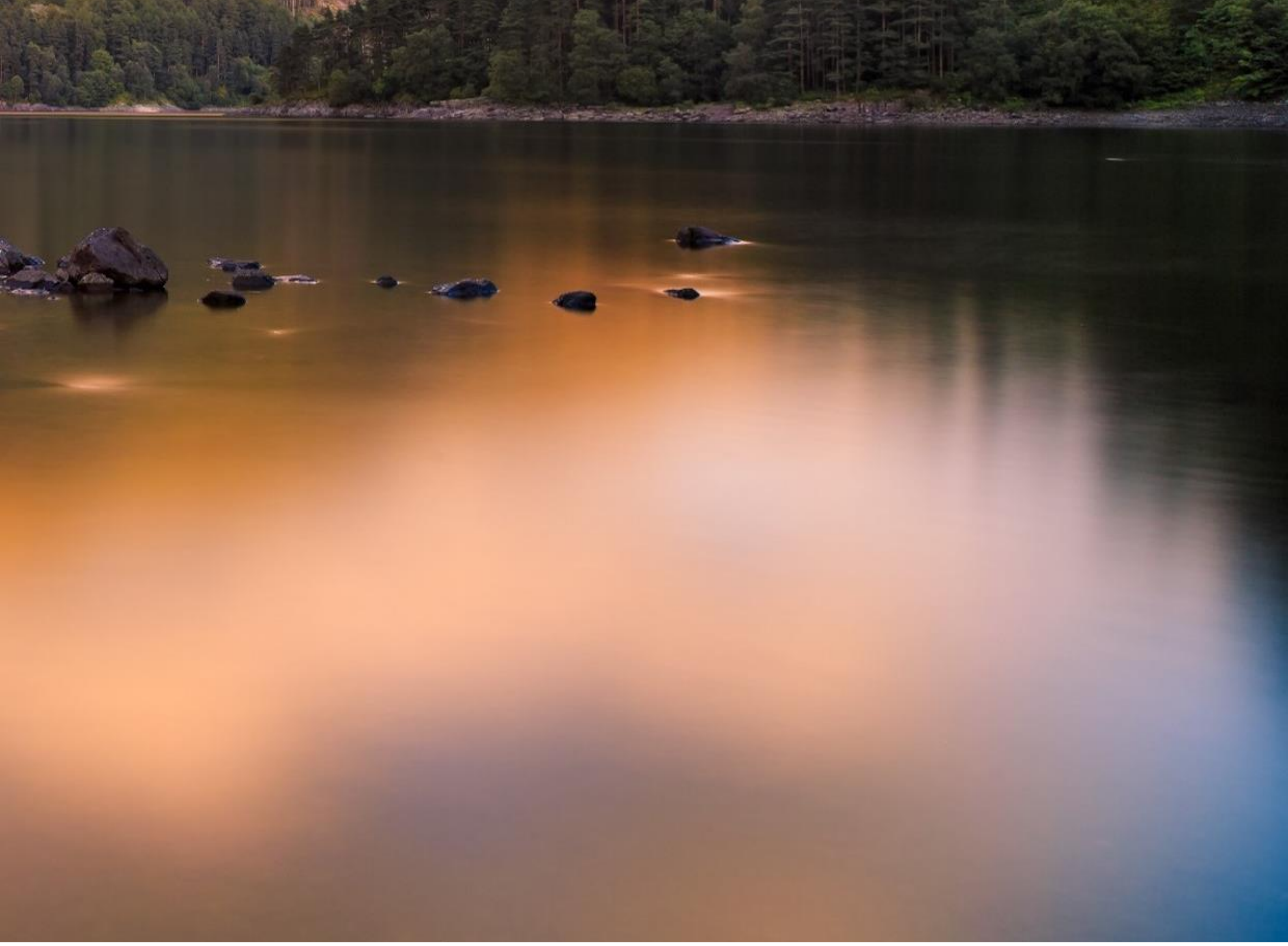
Future Ideas Lab

What lessons can we learn from cost assessment at PR19?

February 2022

Private

In the first paper of our cost assessment-related Future Ideas Lab series, we set out six key principles for regulatory cost assessment. In this paper, we apply these principles to some key areas of the PR19 framework and ask what lessons we can learn for PR24 as a result.



United Utilities

Water for the North West

Executive Summary

Our Future Ideas Lab paper '*The Principles of Regulatory Cost Assessment*'[1] set out six key principles for cost assessment. These principles were based upon our reflection of cost assessment at PR19 and a look ahead to the challenges the sector will face at PR24. In this paper, we apply these principles to some key areas of Ofwat's cost assessment framework at PR19 and ask what lessons we can learn for PR24 as a result.

Ofwat's approach to cost assessment at PR19 - particularly its approach to developing its model suite - was a significant improvement on that implemented at PR14. Whilst there are opportunities for further improvements, we recognise that Ofwat got a lot of things right with its approach. We are encouraged by Ofwat's continued engagement with the industry on cost assessment and consider that its PR19 approach will provide a strong foundation for PR24.

The aim of this paper is explore key areas of Ofwat's methodology at PR19 through the lens of our '*Principles of Cost Assessment*' paper[1]. We have focused on a small number of key areas where we consider that there is scope for the framework to be improved at PR24. These areas are:

- The cost-service relationship;
- The approach to implicit productivity challenges; and
- Ensuring the approach is sufficiently diverse to capture the interactions between cost and cost driver for each service provided by each company.

This paper explores whether Ofwat's approach in these areas was consistent with our principles of cost assessment. It then considers what action could be taken in future price reviews to better align its approach to cost assessment with these principles.

Exploring the relationship between cost and service

Our *'Principles of Cost Assessment'*[2] paper argued that cost assessment must be led by engineering, operational and economic rationale. While we are sympathetic to Ofwat's view that companies have been seen to deliver good outcomes performance and cost efficiency simultaneously[3], we consider that engineering, operational and economic rationale supports the existence of a cost-service relationship. An often overlooked, but key element of the cost-service relationship is that the relationship can change depending upon the operating circumstances of a particular region. In such cases, a common target will create inequitable stretch across the industry, with some companies receiving a relatively achievable target while for others the same target may be unattainable. The implication of this is that some customers are paying too much for the service they receive.

At PR19, Ofwat set a costs and outcomes package that it believed was stretching but achievable 'in-the-round'. This was generally supported by the Competition and Markets Authority (CMA), although the CMA did intervene in some key areas. Ofwat considered a common industry target to be appropriate for water supply interruptions, pollution incidents and internal sewer flooding because: *"…There is no clear reason why companies should not be achieving the same stretching level of performance"*[4] for these performance commitments. This approach was largely accepted by the CMA. Ofwat's 2022 base cost assessment consultation suggested that it does not intend to change its approach in this area for PR24.

In our view, expecting that companies in diverse operating areas can equally attain the same level of performance was a simplifying assumption on Ofwat's part which was often not informed by engineering, operational or economic assessments. There may be times where this simplifying assumption is not unreasonable and does not skew incentives significantly. However, when engineering, operational or economic assessments point to clear

[1] UUW (2021) *The Principles of Regulatory Cost Assessment*. Available here.

[2] UUW (2021) *The Principles of Regulatory Cost Assessment*. Available here.

[3] Ofwat (2020) *Response to cross-cutting issues in companies' statement of case.* Available here.

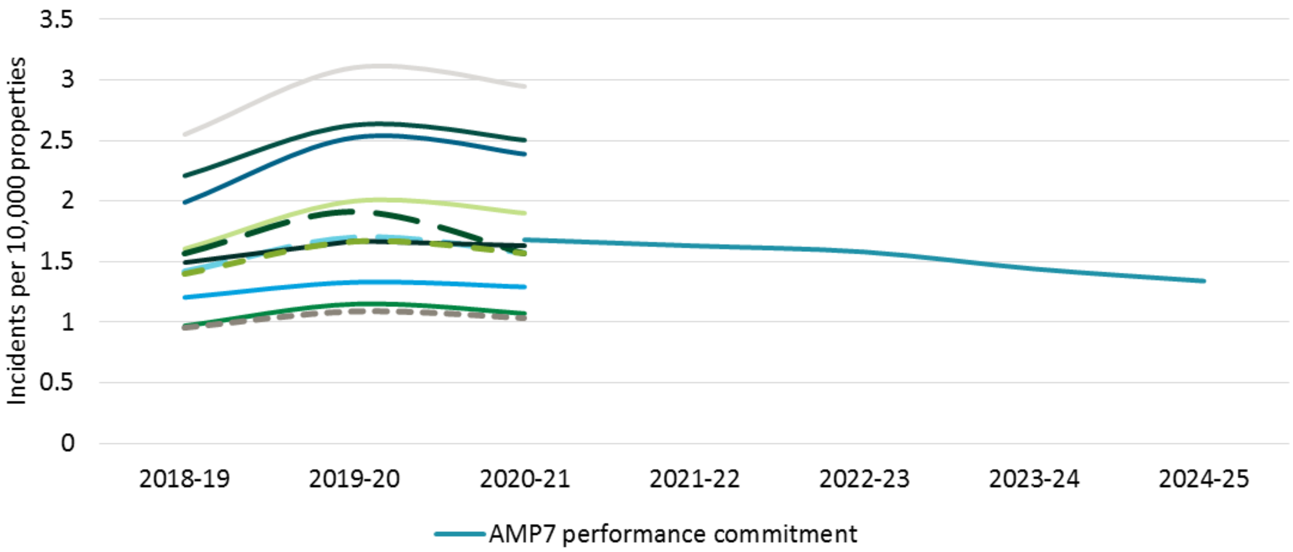
[4] Ofwat (2017) *Delivering Water 2020: Our final methodology for the 2019 price review.* Available here.

reasons why performance should be expected to be different, then a uniform approach would distort incentives between companies, leading to suboptimal outcomes for customers. If the framework serves to unduly penalise some companies while improperly benefiting others then customers in certain areas will receive an unreasonably expensive service as a result.

Therefore, we have developed a framework that assesses whether a common industry target or a company-specific target is more appropriate. This framework considers the key performance drivers for each measure and – in particular - assesses whether these drivers are under management control or are clearly external factors imposed by the operating environment. We use this framework to set a company-specific target for internal sewer flooding[5], using a 'proof of concept' econometric modelling approach analogous to the way that Ofwat sets cost targets. The resulting company-specific target can be compared to the common industry target to evidence whether the common target is appropriate. If it is not, the company-specific target calculated by the performance model can be implemented. The framework is illustrated in Figure 2Figure 2 - Our framework for assessing whether a company-specific target is more appropriate than a common industry target on page 4.

Figure 1 sets out company-specific upper quartile levels of performance calculated by our models, depending upon the characteristics of each company's region. It's clear that the models expect regional characteristics to create a wide variety of performance, and that setting a common industry target may be unrealistic and could lead to customers in some areas paying too much for the service they receive.

Figure 1 - Modelled upper quartile performance levels compared to the common AMP7 performance commitment



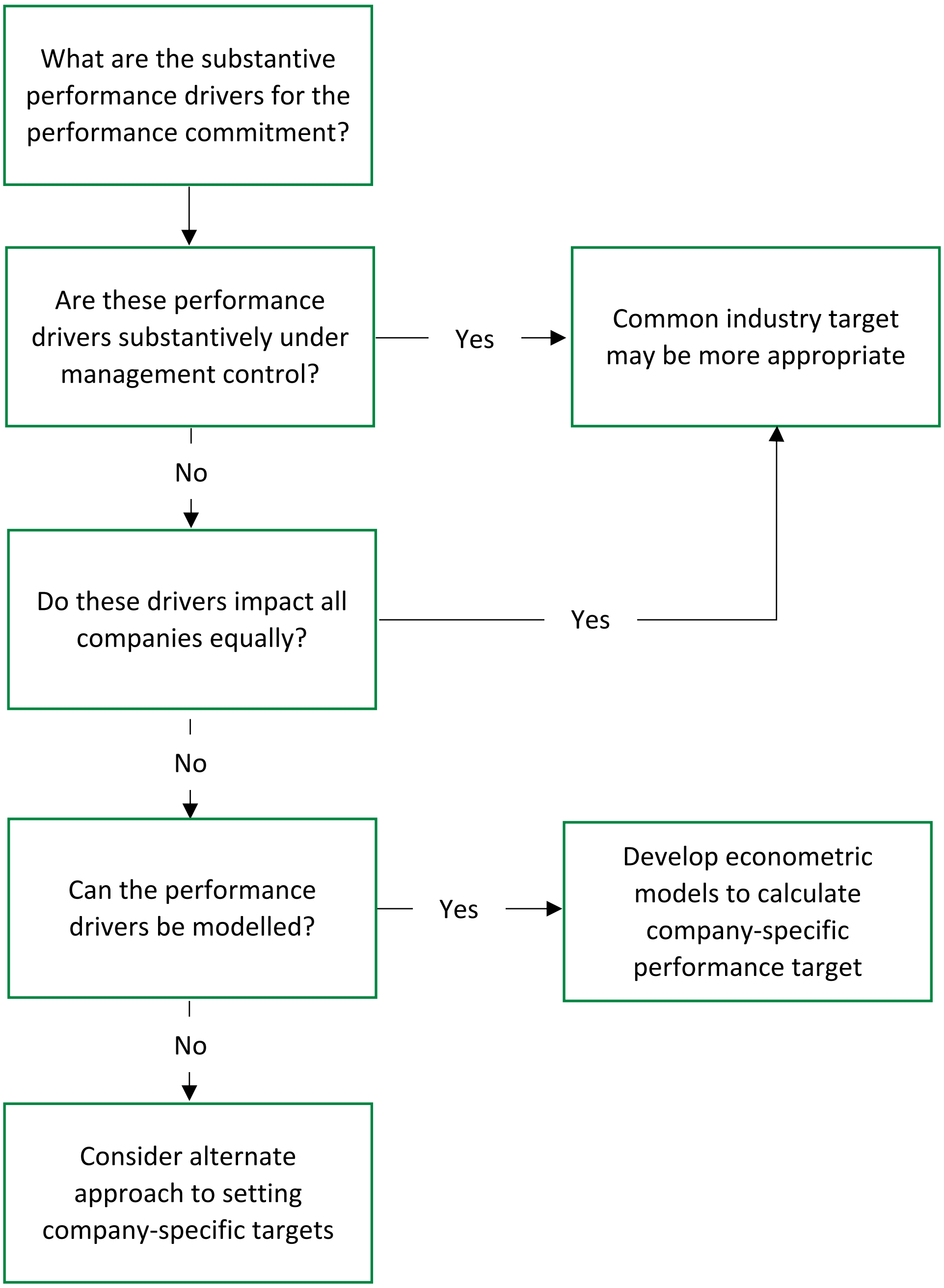
We consider that company-specific targets produce a more equitable position for companies across the industry, leading to better incentivisation and more optimal allocation of resources for the benefit of customers. In the case of sewer flooding companies operating in a fairly benign area for internal sewer flooding performance receive a challenge that is as equally stretching as companies operating in more challenging environments, meaning that customers across the country are paying for an equivalently efficient and stretching level of service.

This approach represents a departure from the PR19 framework where Ofwat assumed that the cost assessment process would be able to account for differences in operating conditions to the extent that a common target becomes appropriate. However, this assumption would only hold true if the variations in required costs were small and targets were equally achievable by all companies. We believe that at PR19 this was assumed but not proven. As an alternative, we are promoting that, in some circumstances whereby targets are not equally achievable, econometric models can be used to predict an equally efficient level of performance for each

[5] We use internal sewer flooding as a specific example, but we note that the approach could be applied to other performance commitments.

company, which should then be achievable for all companies from a common assessment of costs. This means that all customers pay for an equivalently stretching level of service.

Figure 2 - Our framework for assessing whether a company-specific target is more appropriate than a common industry target



Challenging efficiency within a transparent, stable and objective framework

The previous section explored how the PR19 approach to the cost-service relationship can lead to inequitable outcomes for customers across the industry, and suggested a specific solution. This section considers how the PR19 cost assessment framework created implicit productivity challenges more generally.

Regulators mimic the effect of competition by using comparisons between companies to challenge companies to reduce cost requirements and improve service performance. Efficiency assessment presents a number of important issues to consider, and careful thought needs to be given to designing and maintaining a transparent, objective and stable framework if the aim is to produce an outcome that is in the best long-term interests of customers. We consider that greater transparency is needed to allow companies (and customers, shareholders and other stakeholders) to fully understand the scale of efficiency challenges that companies are being assumed to meet.

The key area that we consider would have benefitted from additional transparency was the overall efficiency stretch being applied to companies – in particular by comparison with previous AMP periods. Ofwat applied stretch across the framework, but without an explicit, holistic assessment of the overall extent of the challenge. We do note that Ofwat judged that the settlements were deliverable 'in-the-round' but we are not aware of any analysis that accounted for all the different areas of stretch, both explicit and implicit (we note that Ofwat provided some analysis in its submissions to the CMA[6] but this did not account of all the implicit challenges we set out below). We set the different sources of stretch out below, both explicit and implicit. If the PR19 framework accounted for all these elements in its assessment of stretch, this would have provided added confidence that Ofwat's determinations fully reflected reasonable expectations for company efficiency.

Explicit areas of stretch

- Ofwat applied a 'catch-up' efficiency challenge, by reference to a benchmark company. At PR19, this benchmark was the fourth most efficient company in water and the third most efficient in wastewater, although the Competition and Markets Authority (CMA) reduced this challenge to the upper quartile level. This represented a 4.6% improvement in efficiency in water and 2% in wastewater, compared with the average level of efficiency in the sector (albeit the actual efficiency requirement varied by each company)
- Ofwat applied a 'frontier-shift' challenge, which applied an ongoing productivity challenge to companies of 1.1% per annum, relative to its newly implemented inflation measure of CPIH.

Implicit areas of stretch

- The move from an RPI-linked control to a CPIH-linked control automatically imparted a circa 1% per annum additional challenge on companies (compared with historic frontier shift assumptions) due to the wedge between the two indices;
- Inflation reflects economy-wide productivity gains. Therefore if Ofwat assumed zero (relative to CPIH) for its dynamic efficiency assumption, it would be incorrect to conclude that no dynamic efficiency is assumed – it is just that no additional dynamic efficiency is assumed over and above that being delivered by the economy as a whole;
- Retail price controls are in nominal terms, meaning that inflation pressures were implicitly assumed to be absorbed by companies - this represented an additional ongoing productivity challenge, over and above the explicit forward-looking challenge applied;
- Pursuing a modelling methodology of sensible simplicity while maintaining a high evidential bar for cost adjustment claims reduces the likelihood that companies' benchmarks reflect the challenges their operating regions pose, effectively constituting an additional 'catch-up' efficiency challenge; and
- Ofwat's requirement that companies must achieve the delivery of more stretching performance through base costs (for example, the 15% leakage reduction[7]) means that companies were required to achieve an additional frontier shift efficiency (to absorb the cost of meeting those higher performance levels) without a further adjustment to the baseline. The actual frontier shift required will differ from company to company depending on their changes in performance and regional operating conditions, but it was significantly greater than zero for each company. A coherent framework should not overlook this when making assumptions about the forward-looking efficiency challenge.

At PR24, we consider that a coherent framework should take all these sources of stretch into account particularly given the scale of the challenges the sector is facing.

---

[6] Ofwat (2020) *Introduction and overall stretch on costs and outcomes*. Available here.

[7] Wholesale Water Supply-demand balance enhancement – feeder model summary p12, Ofwat

A diverse model suite better reflects the costs of delivering water and wastewater services

It is unrealistic to expect that any one model or any suite of models is capable of perfectly capturing the relationship between cost and cost drivers across the industry with the data available. The heterogeneity of operating conditions across the water industry means that a more diverse approach is much more capable of effectively recognising efficient costs for each company. This is because:

- It maximises the information that is used to build the benchmark, which should increase the engineering, operational and economic rationale underpinning the cost assessment framework;
- There is more potential to capture key cost drivers across the different services that companies provide through the same value chain, or capture the same cost driver in an alternative way, using a different variable. This should reduce the risk of the choice of variable unduly benefiting/penalising certain companies;
- The need for cost adjustments decreases. This places more weight on the modelled benchmark, which is a well-understood and transparent assessment method. This should help to reduce problems relating to information asymmetries between regulator and companies. It also means that both regulator and companies realise substantial time and cost savings related to evidencing and assessing cost adjustment claims; and
- It allows different aggregations of the value chain to be included, which will provide a better understanding of substitution effects across the value chain within the benchmark.

At PR19, Ofwat made a policy decision to pursue a ‘sensibly simple’ model suite. It identified four key cost drivers (scale, complexity, density and topography) and sought to reflect these in its models using a parsimonious selection of cost drivers. Sensible simplicity may be part of a legitimate methodology, but only if it recognises that it increases the likelihood that companies with particularly adverse operating conditions would be significantly penalised absent a proportionate assessment of cost adjustment claims, particularly given the complex and variable operating conditions faced by companies (particularly in wastewater) and the relatively small dataset available (which means that any single model is likely to be inherently weak and prone to unreliability). The sensitivity to these issues is further exacerbated by the level of stretch Ofwat imposed elsewhere in its framework. A coherent framework should recognise the relationship between the modelling methodology and the subsequent impact upon the cost adjustment assessment process.

However, we do consider that Ofwat adopted an appropriately diverse set of cost aggregations including water resources plus and bioresources plus. These cost aggregations recognised that companies will structure their assets differently across the price control boundaries, and as a result a benchmark which doesn’t appropriately account for these ‘substitution effects’ is likely to be inappropriate.

At PR24, we consider that Ofwat should continue to implement diverse aggregations of cost that have the potential to capture substitution effects within the benchmark. We consider that Ofwat should augment these diverse cost aggregations with a more diverse basket of cost drivers (utilising triangulation between different model forms) which should better reflect the diversity in in operating conditions across the industry. This should alleviate the need for cost adjustments, increasing the transparency of Ofwat’s assessment and reduce administration burdens on companies and Ofwat.

Ofwat could also implement targeted triangulation across its different models but also across the different cost aggregations. Unlike simple triangulation which implicitly assumes all models are equally capable of predicting costs for every company, targeted triangulation seeks to minimise the gap between predicted and actual costs by weighting a model’s prediction proportionately to its ability to predict costs for each company. This has a symmetrical effect of reducing the risk that companies are unduly benefitted by a model but also that companies are unduly penalised.

Contents

Appendices

1. Introduction

Effective economic regulation helps to ensure that monopolies deliver the service that customers want for the right price. In theory, the absence of competition can reduce incentives for a monopoly to continuously improve efficiency and service performance, compared to companies operating within a competitive market. The role of the regulator is to try to mimic the pressures of competition through its framework and actions. One of the ways in which it achieves this is through the Price Review, where it sets the price, investment and service package that customers should expect to receive from an efficient company.

This requires a delicate balance within the Price Review framework. On the one hand, it has to ensure that companies are able to finance the delivery of water and wastewater services expected by customers and meet their statutory obligations. On the other hand, it must protect the customer interest, and seek to prevent customers paying for inefficient costs.

One of the key issues for the regulator in determining the investment requirements for a company is the asymmetry of information that exists between regulator and company, whereby the regulator does not know the company's efficient level of cost. Because of this, the regulator will tend to create its own independent view of efficient costs through the development of an independent benchmark, whilst also incentivising companies to reveal efficient cost requirements within their business plans.

The framework the regulator uses to achieve these objectives is cost assessment. Cost assessment often utilises economic and engineering logic, mathematical modelling and statistical analyses to arrive at an independent benchmark for a company. The regulator can design a framework to encourage companies to reveal efficiencies, and can inform its benchmark with engineering narratives to ensure that it reflects the reality of the company's operations. Crucially, the benchmark's independence ensures that the company faces incentives to outperform, benefitting customers in the long run. In this way, a well-designed framework can achieve the delicate balance required of utility regulation, mimic a competitive market and deliver a great outcome for customers.

Ofwat's framework at PR19

Ofwat outlined through its methodology that PR19 would require a step change for the sector. It involved a disaggregation of the wholesale price controls to facilitate the development of competition within bioresources and water resources markets. It aimed to deliver greater focus on the shared strategic objectives of each company in delivering resilience for the wider economy. It encouraged companies to seek new opportunities to innovate to deliver better services for their customers. It was in this context that Ofwat designed its cost assessment framework.

The framework removed the menu approach to cost sharing, replacing it with a simpler approach that incentivised companies to submit efficient business plans. Ofwat did not utilise overly complicated econometric models to assess total expenditure (totex) as at PR14, instead assessing most enhancement spend separately and focussing its 'sensibly simple' approach to econometric modelling only on base costs (although it included some growth enhancement costs within base models at the Draft Determinations stage). Ofwat assessed wastewater environmental improvements 'in-the-round', and made greater use of engineering assessments for smaller value enhancements, replacing the unmodelled uplift from PR14 with deep dives of enhancement business cases.

We consider this framework, particularly its approach to developing its model suite, to be a significant improvement on that used during PR14. Whilst there are opportunities for further improvements, we recognise that Ofwat got a lot of things right with its approach to cost assessment at PR19. For the most part, Ofwat developed its framework transparently and through consultation with industry stakeholders. This involved significant early engagement with the industry, through various working groups and consultations. Indeed, our own interpretation of the Competition and Markets Authority's (CMA) PR19 redetermination outcome is that it largely supported Ofwat's approach to cost assessment. This is in contrast to the PR14 redetermination for Bristol Water, where the CMA was more critical of Ofwat's framework and instead chose to develop its own approach.

The purpose of this paper

The aim of this paper is to explore key areas of Ofwat's methodology at PR19 through the lens of our *'Principles of Cost Assessment'* Future Ideas Lab paper[8]. We consider Ofwat's PR19 approach to be a substantial improvement on its PR14 approach. However, there is some scope in key areas for the framework to be improved at PR24. These areas are:

- The cost-service relationship;
- The approach to implicit productivity challenges; and
- Ensuring the approach is sufficiently diverse to capture the interactions between cost and cost driver for each service provided by each company.

This paper explores whether Ofwat's approach in these areas was consistent with our principles of cost assessment. It then considers what action could be taken in future price reviews to better align its approach to cost assessment with these principles.

[8] UUW (2021) *The Principles of Regulatory Cost Assessment*. Available here.

2. Exploring the relationship between cost and service

Intuitively, higher service quality should be associated with higher costs. This relationship can be observed in most competitive markets. Therefore, it may be unrealistic to expect this relationship not to hold in the water industry. Such an expectation creates a risk that setting performance targets independently of cost targets could create an unachievable frontier. The PR19 approach frequently asked companies to fund service improvements from base, without explicitly factoring the scale of such challenges into its cost baseline. If there are exogenous factors that significantly contribute to company performance in any one area, then this approach will have an unequitable impact across the industry, meaning customers in some areas are paying too much for the level of service they receive.

Therefore, we consider that there is considerable scope to improve the price review framework by considering whether the cost-service relationship could vary across the different regions of England and Wales. In such cases, it may be appropriate for Ofwat to implement company-specific targets or adjust costs to acknowledge the greater stretch imposed on companies operating in challenging regions.

2.1 Engineering, operational and economic rationale of the cost-service relationship

Our *'Principles of Cost Assessment'*[9] stated that cost assessment must be led by engineering, operational and economic rationale. In this section, we apply this rationale to the cost-service relationship.

Economic rationale supports the cost-service relationship. Competitive markets provide consumers with a diverse range of choice. One of these choices is service quality. Consumers recognise that higher quality service tends to mean higher prices and are able to choose a combination of price and quality that aligns to their preferences. A profit-maximising company that is able to provide the same quality product at a lower cost would win the consumer's business, so the fact that different price points exist suggest higher quality products generally lead to higher costs otherwise companies would cut prices to maximise market share. This economic logic suggests that a higher quality product is associated with higher costs such that companies coalesce around a cost-service trade-off frontier.

Engineering rationale also suggests that higher service quality is associated with higher cost. One example from the water sector is in wastewater treatment, where additional stages of treatment must be implemented to improve the quality of the final effluent. These stages of treatment are associated with higher capital costs as well as higher ongoing operational expenditure like higher power costs due to UV treatment processes for example. While technological advances can reduce costs, these aren't always viable solutions and long-lived assets mean that companies will still need to operate solutions implemented before the new technology was available.

Operational rationale suggests that higher service quality may be associated with higher costs. For example, the risk of sewer blockages can be minimised through regular sewer flushes, which will increase operational expenditure. However, it may be possible for companies to optimise their activity by developing a better understanding of their networks.

A key feature of the cost-service relationship is that the relationship may not be the same for all companies. Engineering and operational logic dictate that the operating circumstances of a region can present particular challenges. While some of these regional effects can be mitigated through appropriate management action, this is not always possible. If this (or indeed the wider cost-service relationship) is not accounted for by the regulator, then the presence of these regional effects will mean that the level of stretch imparted by the regulatory settlement is unequally distributed across the industry, meaning that customers in certain areas will pay too much for the level of service they receive.

[9] UUW (2021) *The Principles of Regulatory Cost Assessment*. Available here.

We are sympathetic to Ofwat's view that companies have been seen to deliver good outcomes performance and cost efficiency simultaneously[10], and that therefore additional stretch is appropriate. However, we note that alternate conclusions can be drawn from this observation. In our view, regional operating conditions can have a material impact upon both companies' costs and performance. Where these circumstances are not reflected in the botex models or the approach to setting performance commitments, companies operating in areas with favourable characteristics could be seen to perform well on both costs and services. This is because a common approach across the industry is placing additional stretch on some companies with an adverse operating environment while reducing it for others that operate in a more benign environment.

Ofwat may also have implicitly assumed that such differences would be accommodated within cost assessment, e.g. if it might cost companies different amounts to meet a common target level of performance. This assumption would only hold true if the variation in required costs were small, and targets were equally achievable by all companies, which was assumed but not proven. Alternatively, we are promoting that, in some circumstances whereby targets are not equally achievable, that econometric models can be used to predict an equally efficient level of performance for each company, which should then be achievable for all companies from a common assessment of costs.

2.2 Ensuring that the approach to the cost-service relationship is coherent

In our *'Principles of Cost Assessment'*[11] paper, we advocated for the cost assessment framework to be coherent. Coherency in this case requires that the approach to cost assessment is joined up with the approach to setting performance targets.

In areas where engineering, operational and economic logic suggests that cost-service relationship exists, a coherent framework must reflect this in cost assessment or make offsetting adjustments to performance targets. Otherwise, the package as a whole will be unachievable. Additionally, where engineering, operational and economic rationale holds that regional operating conditions impact upon performance, the framework should take this into account. Otherwise, some companies will receive additional challenge and customers of other companies will pay too much for the level of service provided.

2.3 Ofwat's approach at PR19

At PR19, Ofwat considered that it set a costs and outcomes package that 'in-the-round' was stretching but achievable. This was generally supported by the Competition and Markets Authority (CMA), although the CMA did intervene in some key areas. Because the cost-service relationship received significant airtime during the CMA redeterminations, we will concentrate on one area we consider did not receive enough focus during PR19; the approach to setting the targets for the 'upper quartile' performance commitments (water supply interruptions, pollution incidents and internal sewer flooding) and the subsequent implications for cost assessment.

Ofwat asked companies to provide a forecasted upper quartile level of performance for each PC, and then set the PC target at the upper quartile of these estimates (other than water supply interruptions where it implemented a slightly less stretching target). It justified this approach through reference to the information asymmetry that exists between companies and the regulator[12]. Ofwat considered a common industry target to be appropriate because it considered: *"…There is no clear reason why companies should not be achieving the same stretching level of performance"*[13] for these performance commitments. This approach was largely accepted by the CMA. Ofwat's 2022 base cost assessment consultation suggested that it does not intend to change its approach for PR24.

In our view, expecting companies in vastly different operating areas can achieve the same level of performance was an assumption on Ofwat's part. We are not aware of any engineering, operational or economic assessment

---

[10] Ofwat (2020) *Response to cross-cutting issues in companies' statement of case.* Available here.

[11] UUW (2021) *The Principles of Regulatory Cost Assessment*. Available here.

[12] Ofwat (2019) *Delivering outcomes for customers policy appendix.* Available here.

[13] Ofwat (2017) *Delivering Water 2020: Our final methodology for the 2019 price review.* Available here.

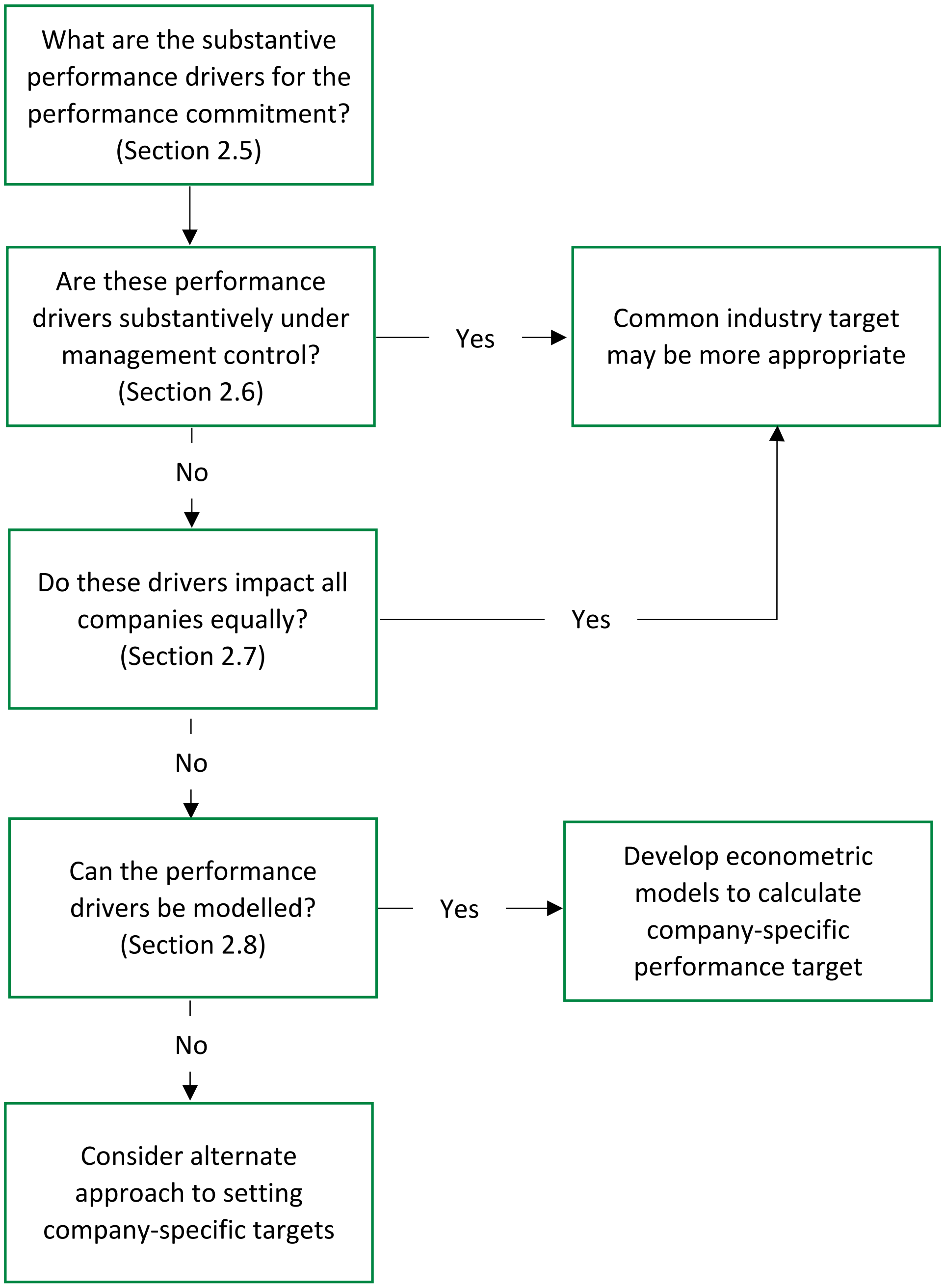
that informed this policy decision. If in fact there are clear reasons why performance can be expected to be different, then this approach would unduly penalise some companies while improperly benefiting others. This means that the level of stretch across the industry is inequitable, and the customers of companies which benefit will pay too much for the level of service received.

In the remainder of Section 2, we develop an approach that explores the engineering, operational and economic rationale underlying the cost-service relationship for the 'upper quartile' PCs and suggests approaches to make appropriate adjustments to ensure that the overall framework is coherent.

2.4 How to account for the cost-service relationship within a coherent framework

In its recent base costs consultation, Ofwat suggested that it would continue to implement a common target for water supply interruptions, pollution incidents and internal sewer flooding. In the following sections, we suggest a framework to assess whether performance commitments should be company-specific or common across the industry. We then develop an approach to calculate company-specific performance targets using the specific case of internal sewer flooding. This company-specific target can then be compared against the common target to evidence which target is most appropriate. Our approach is illustrated in Figure 3. This approach could be applied to any performance commitment, not just the upper quartile group at PR19.

Figure 3 - Our framework for developing performance targets



2.5 What are the substantive engineering, operational and economic performance drivers for the performance commitment?

This section considers the engineering, operational and economic rationale underlying the performance measures Ofwat currently subjects to a common target.

Figure 4 - What are the key drivers of water supply interruptions performance?

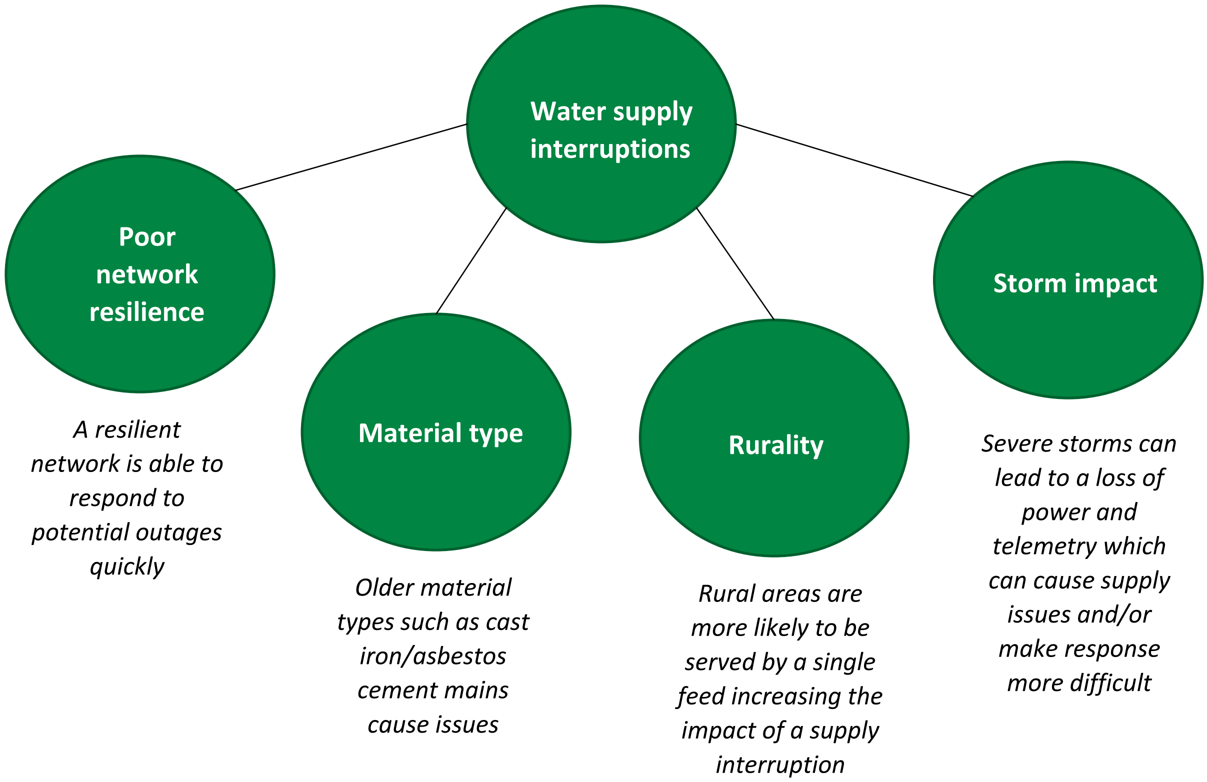


Figure 5 - What are the key drivers of pollution incidents performance?

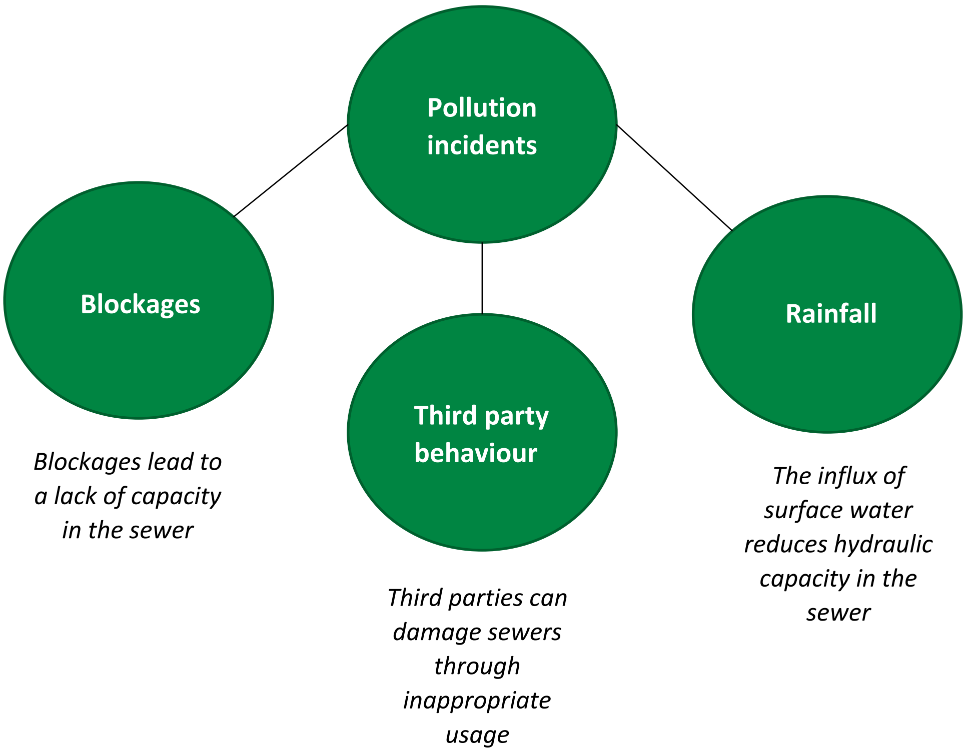
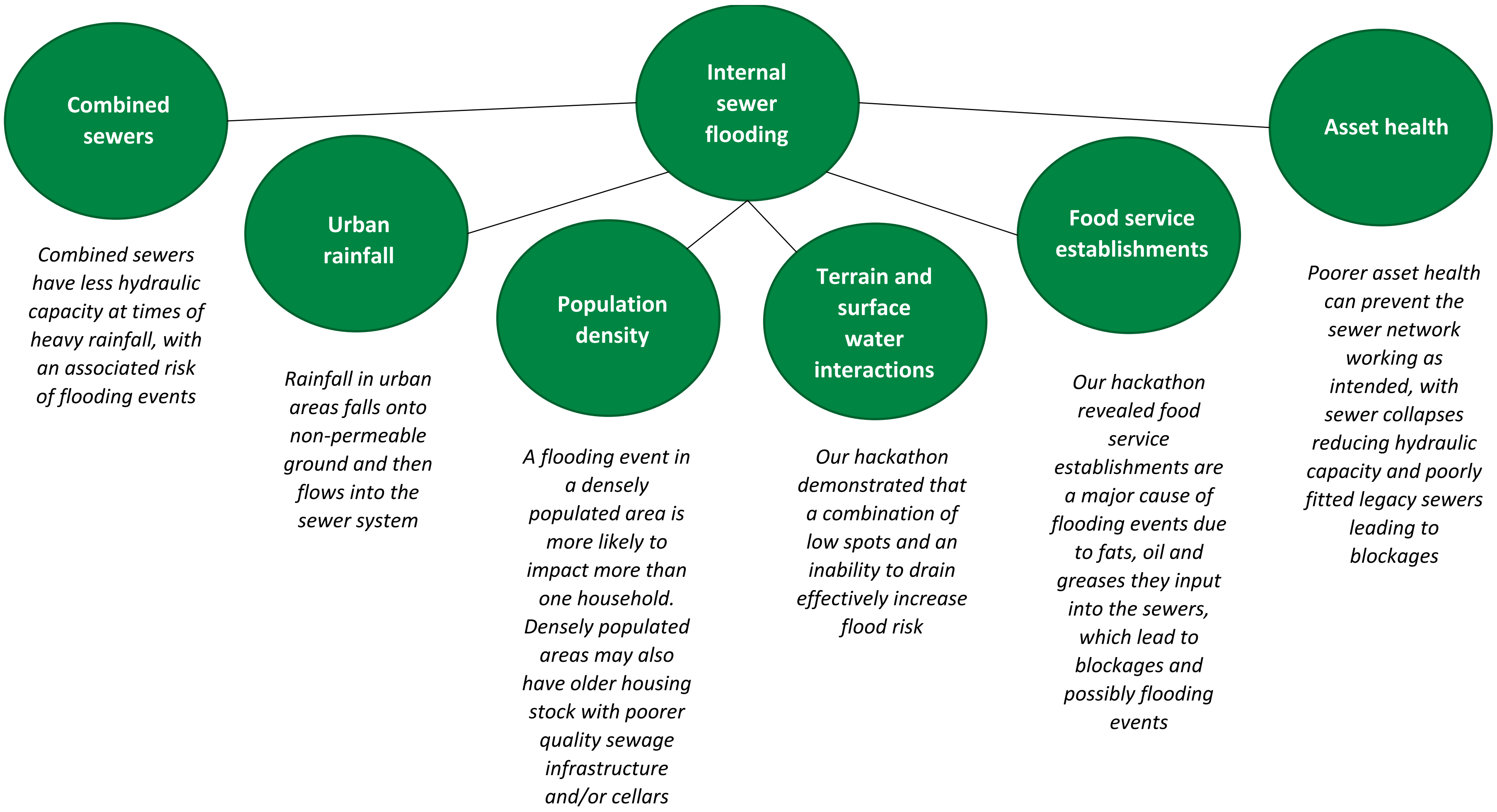


Figure 6 – What are the key drivers of internal sewer flooding performance?*



*This assessment was informed by our internal sewer flooding 'hackathon', where we drew upon a wide range of data to assess the key drivers of flooding performance in our region. We summarise the hackathon in more detail in Section A.4.

2.6 Are these performance drivers substantively under management control?

In section 2.5, we set out the engineering, operational and economic rationale that help us to understand variations in company performance. This section considers whether these 'performance drivers' are substantively under the control of company management or whether they are 'exogenous', meaning that they are outside of the control of company management. Focusing upon exogeneity ensures that performance targets reflect the difficulties imposed by local operating conditions, but do not reflect differences in management practices.

Table 1 - Is water supply interruptions performance materially driven by exogenous factors?

Performance driver	Under management control?
Poor network resilience	Management is able to: • Invest in centralised monitoring and control to detect problems early and appropriately coordinate any potential response. • Invest in means of alternate supply like strategic water mains and mobile water tankers. While the planning of strategic water mains is a longer-term ambition, we consider that it is possible to establish an effective water tanker fleet in a relatively short time-frame. • Develop appropriate contingency plans to ensure continued supply or rapid return of supply during incidents. • Ensure staff have appropriate training to ensure the network is operated and maintained in an optimal way, for example through the use of 'calm networks'.
Material type	The material type of pre-privatisation water mains is outside of management control. While materials such as cast iron and asbestos cement pipes are problematic and driving asset health issues, it is likely to impact on all companies to an extent.
Rurality	While the rurality of an area is outside of management control, companies are able to manage its risk and impact through improved monitoring or by optimising the location of their mobile water tankers.
Storm events	Severe weather events are outside of management control, but companies are able to mitigate their impact upon telemetry and water supply through the use of back-up generators and mobile water tankers.

Therefore, we consider that exogenous factors are **not a material driver** of performance and that it is reasonable for this performance commitment to be subject to a common industry target.

Table 2 - Is pollution incidents performance materially driven by exogenous factors?

Performance driver	Under management control?
Blockages	The impact of blockages can be mitigated through appropriate operating practices. For example, management is able to monitor capacity in the sewers and optimise the response to any emerging problems. Management can also ensure the network is properly maintained and flushed to maximise capacity. Heavy rainfall can make the impact of blockages worse.

Third party behaviour	Isolated incidents relating to customer or third party behaviour are out of the control of companies, but companies are able to respond to systematic issues through appropriate engagement with relevant parties.
Urban rainfall	Rainfall is outside of management control. However, rainfall produces much less concentrated effluent meaning that any spill is more likely to be classed as a Category 4 event. These are not currently covered by the pollution incidents performance commitment. Therefore, we think it's reasonable to exclude rainfall from our consideration of whether pollution incidents (as defined in the performance commitment) is under the substantive control of management.

Therefore, on balance we consider that **exogenous factors are not a material driver** of performance as defined by the **current definition** of the performance commitment, and that it is reasonable for this performance commitment as currently defined to be subject to a common industry target.

We note that this assessment **may change** if the regulatory framework surrounding pollution incidents changes. For example, if the definition of the performance commitment changes to include factors outside of management control it may become appropriate to consider whether company-specific targets are suitable.

Table 3 - Is internal sewer flooding performance materially driven by exogenous factors?

Performance driver	Under management control?
Combined sewers	Combined sewers are a legacy asset, inherited by companies at privatisation. While companies are able to create separate sewer systems, this process is expensive and therefore a long-term ambition.
Urban rainfall	Urban run-off is dictated by rainfall and urbanisation within a region, both of which are outside of management control.
Population density	Population density is outside of management control. Management does have some capacity to enter into partnership schemes to increase the prevalence of blue-green infrastructure, although implementing enough schemes to make a regional-wide impact is a longer-term ambition.
Terrain and surface water interactions	Topography is outside of management control, although we reflect this performance driver using an asset-based measure, which is potentially under management control.
Food service establishments	Outside of management control, although companies are able to engage with FSEs to mitigate any adverse impacts. On balance, we consider it's appropriate to include FSEs in our assessment.
Asset health	As an asset-based measure, asset health is potentially under the influence of company management in the long-term. However we consider the potential for substantive short-term influence to be negligible. Additionally, we note that issues that pre-existed privatisation such as poor-quality materials and poor installation practices continue to impact on company performance. Due to the limited ways of measuring asset health (currently), we use sewer collapses as it is less susceptible to management influence in the short-term.

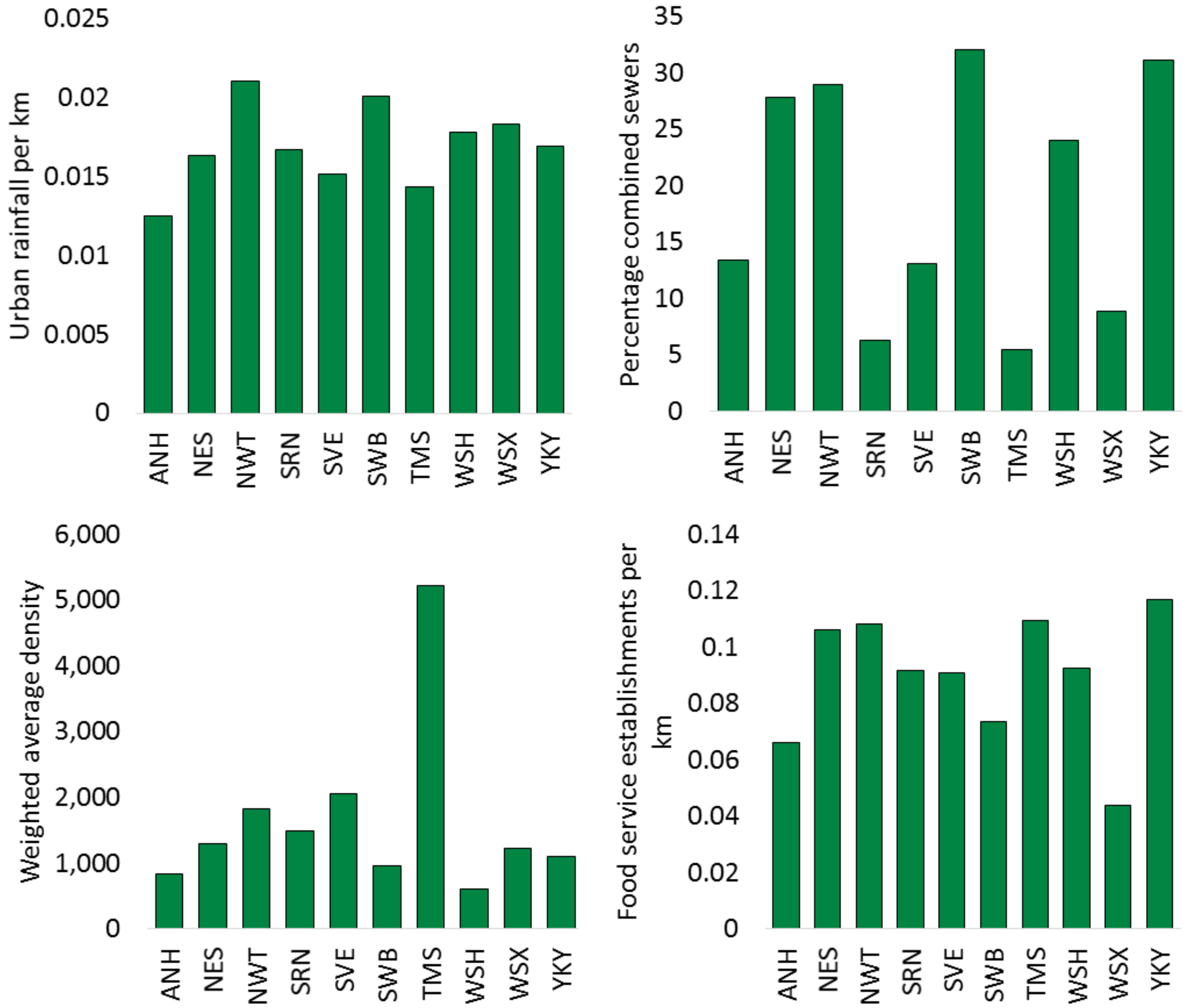
Therefore, overall we consider **that exogenous drivers materially affect internal sewer flooding performance** and that it is reasonable for this performance commitment to be subject to a company-specific industry targets of equivalently stretching performance, reflecting company-specific circumstances.

2.7 Do these performance drivers affect all companies equally?

If exogenous performance drivers impact upon all companies equally then the base cost allowance should provide an appropriate allocation of costs across the industry. If this is the case then a common performance commitment could be appropriate.

Simple analysis suggests that there is considerable variation across the industry for the key performance drivers of internal sewer flooding identified in the previous section. Therefore, it is appropriate to seek to develop company-specific performance targets for this PC.

Figure 7 - There is substantial industry variation in key internal sewer flooding performance drivers



2.8 Can these performance drivers be modelled?

Econometric models provide a transparent way to assess whether the exogenous performance drivers identified through our engineering, operational and economic assessment impact upon company performance across the industry. A main constraint of implementing an econometric approach is the availability of consistent data which reflects the key performance drivers. In this case, we have identified robust sources of data for most of the main performance drivers of internal sewer flooding meaning we are able to develop econometric models.

In some cases, such data may not be readily available meaning Ofwat may have to consider alternative approaches to setting company-specific targets. However, this is outside the scope of this paper so we do not consider this further here.

2.9 Develop econometric models to calculate company-specific performance target

Econometric models allow us to directly test Ofwat's PR19 assumption that it: *"did not consider there to be clear reasons why companies should not achieve the same stretching levels of performance for these PCs"*[14]. We are able to include exogenous performance drivers within an econometric model and use the associated framework of statistical significance to conclude whether the evidence supports Ofwat's assumption that these exogenous factors do not drive internal sewer flooding performance.

In order to promote transparency, we have generally followed Ofwat's methodology and approach to developing econometric cost models. Where we considered that an alternative approach was more suitable in this context, we provide clear justification.

2.9.1 Our approach to modelling performance is aligned to Ofwat's approach to modelling costs

Ofwat should be commended for the degree of transparency it provided throughout the PR19 process. Ofwat engaged with the industry very early through working groups and its methodology consultation. It consulted on its approach to cost modelling and provided detailed commentary explaining its decisions at each stage (Initial Assessment of Plans, Draft Determination and the Final Determination). This commentary included the means for companies to replicate Ofwat's assessment, with Stata code and associated Excel models.

This transparent approach has allowed our flooding performance modelling methodology to be guided by Ofwat's framework. As a result, we consider that the analysis will be understandable and transparent for all stakeholders.

2.9.2 The interaction between combined sewers and urban run-off

There is strong engineering, operational and economic rationale to suggest that the interaction between combined sewers and urban rainfall is a key driver of performance. However, in a model where these factors are included as separate independent variables, the interaction effect is ignored because econometric models estimate the marginal effect of each independent variable holding all other variables constant. This is a problem because it would mean the model effectively ignores the interaction between combined sewers and urban rainfall. For this reason, we include an interaction term that captures the inter-relationship between combined sewers, urban rainfall and internal sewer flooding performance.

The interaction term multiplies the percentage of combined sewers by urban rainfall, with each independent variable also included. Note that we have chosen to use total urban rainfall in the interaction term, rather than urban rainfall normalised by sewer length. This approach avoids a squared denominator in the interaction term, produces intuitive results and still reflects the interaction between urban rainfall and combined sewers.

This use of an interaction term adds some complexity, but we consider that the benefits of stronger engineering and operational rationale outweigh the costs of a slightly more complex modelling approach. We consider this approach is still sufficiently transparent to all stakeholders. We provide full details of how to interpret models with an interaction term in section A.2.

2.9.3 A transparent approach to setting performance benchmarks

Our approach to estimating performance targets aligns with the approach Ofwat takes to construct cost benchmarks. It would be straightforward to supplement the performance models we develop in this paper with other elements of Ofwat's cost benchmarking approach. For example, by forecasting performance drivers to calculate future performance targets; by applying a catch-up 'efficiency' challenge; and by adding a frontier shift to reflect technological improvements.

We consider that this means stakeholders will be able to readily understand the approach to setting performance targets.

[14] Ofwat (2021) *Assessing base costs at PR24.* Available here.

2.9.4 Data description and sources

The majority of our data has been sourced from companies' annual performance report submissions. We supplemented this data with the urban rainfall variable created by Arup and Vivid Economics, which we are happy to share with any interested parties.

We have data covering the period 2018-19 to 2020-21 with eleven companies in the panel dataset. However, we have 32 observations, because we were not able to source Thames' internal sewer flooding performance in 2018-19.

2.9.5 Model specification and estimation

We focus primarily on developing a unit performance model where the dependent variable aligns with Ofwat's definition of the performance commitment of incidents per 10,000 properties. It is also a well understood model specification within the industry, as it formed the basis of Ofwat's retail cost models at PR19.

As we noted in section 2.7, we include an interaction term to test the interrelationship between combined sewers, urban run-off and flooding performance. The interaction term is included as an independent variable alongside its individual components. We provide more details about the interaction term and how to interpret it in section A.2.

Ofwat used a random effects estimator in its base cost assessment at PR19 to account for company-specific effects. However, our view is that an OLS estimator is more appropriate in this context. This is because we would like to keep the approach as simple and transparent as possible and OLS is a widely understood estimator. We implement a cluster-robust OLS approach. We note that the models' conclusions are robust to changes in the estimation approach.

We take the natural logarithm of all variables, other than those expressed as percentages. This aligns with Ofwat's PR19 approach and makes it easier for stakeholders to interpret the model results.

2.9.6 Our chosen basket of independent variables

We provide more detail on the independent variables we used in A.1.

2.9.7 Model results

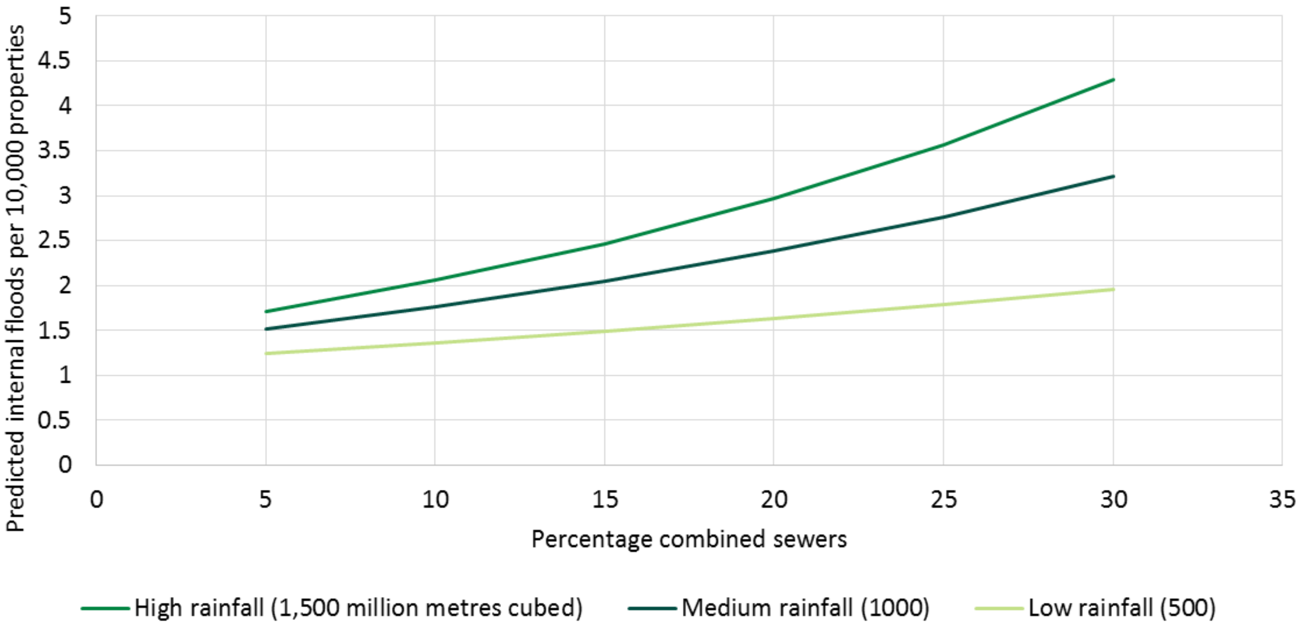
Table 4 presents the model results. We have included asterisks to denote statistical significance and have also included standard errors underneath each coefficient.

Statistical insignificance is not unexpected in a sample size this small. Our key consideration is whether the coefficients' signs support our pre-existing engineering, operational and economic rationale. Notwithstanding this, we are encouraged by the extent of statistical significance across our model suite.

Note that in models with an interaction term, it can be misleading to interpret the coefficient of each component of the interaction individually. This is because such an interpretation would require the other component of the interaction to be zero which is not always appropriate (we provide full mathematical explanation in section A.2). So for example, in model 6 the coefficient of -0.072 tells us the impact of combined sewers on flooding performance *when urban rainfall is zero*. Clearly, this is an unrealistic assumption; all areas of England experienced some urban rainfall in every year of the panel dataset. The negative coefficient on combined sewers is more than offset by the strength of the positive interaction term.

Figure 8 illustrates the effect of the interaction term, using Model 4's coefficients. It's clear the model expects that for any given quantum of combined sewer, companies with more urban rainfall will experience more flooding incidents. While the coefficient on the interaction term is not statistically significant across all models, it is stable. This result supports the underlying engineering and operational rationale.

Figure 8 - An illustration of the effect of the interaction term in Model 4



Both density variables are of the correct sign across all specifications, although weighted average density loses statistical significance in models with an interaction term. However, we do not consider this to be a major drawback given the relatively small sample size. Each measure captures density in a different way so using both will add useful information to the model suite.

The coefficient on number of pumps is negative, which implies that the more pumps are associated with less internal floods. It isn't clear whether this is capturing the effect of topography, but a negative coefficient is consistent with the idea that more pumping macerates sewage which reduces the risk of blockages.

The coefficient on sewer collapses is unintuitive. It implies that more sewer collapses are associated with less internal floods which doesn't align with our engineering, operational or economic rationale.

Food service establishments appear to have a positive effect on internal floods, which aligns with our hypothesis.

Statistical indicators

The R squared scores across most models are encouraging, especially given the unit performance specification. Most models pass the Ramset RESET test for non-linearity in the error term. While models with an interaction term fail the VIF test for multicollinearity (using the conventional cut-off of 10), we do not find this surprising given the interaction term is correlated with other independent variables. Ofwat accepted failed VIF tests in its PR19 models with a quadratic term for the same reason.

Table 4 Results with normalised dependent variable, ln(internal floods per 10,000 properties)

	1	2	3	4	5	6	7	8	9	10	11	12
Combined sewers %	0.021	0.027*	0.026*	-0.087*	-0.051	-0.072	-0.047	-0.052	-0.048	-0.07	-0.038	-0.058
	{0.151}	{0.065}	{0.062}	{0.087}	{0.470}	{0.198}	{0.367}	{0.248}	{0.418}	{0.147}	{0.378}	{0.164}
ln(Urban rainfall per km)	0.144***	0.545***	0.144***	0.205***	0.472***	0.199***	0.440***	0.302***	0.483**	0.224***	0.283*	0.164**
	{0.004}	{0.000}	{0.000}	{0.003}	{0.007}	{0.001}	{0.006}	{0.003}	{0.021}	{0.000}	{0.099}	{0.015}
ln(Weighted average density)		0.385***			0.274*		0.128		0.25		0.134	
		{0.000}			{0.077}		{0.533}		{0.184}		{0.370}	
ln(10,000 properties per km)			2.074***			1.928***		1.701***		1.733**		0.787
			{0.001}			{0.003}		{0.006}		{0.029}		{0.550}
Combined sewers % x ln(urban rainfall)				0.017**	0.012	0.015*	0.01	0.011	0.012	0.015**	0.008	0.012*
				{0.025}	{0.244}	{0.060}	{0.199}	{0.110}	{0.170}	{0.031}	{0.193}	{0.077}
ln(Number of pumps per km)							-0.407	-0.376				
							{0.382}	{0.131}				
ln(Sewer collapses per km)									-0.228*	-0.163		
									{0.096}	{0.319}		
ln(Food service establishments per km)											0.847**	0.625
											{0.041}	{0.366}
Constant	0.925***	-0.295	12.149***	1.109***	0.187	11.521***	1.98	11.497***	-0.833	9.686**	2.729**	6.891
R squared	0.242	0.411	0.467	0.409	0.472	0.611	0.48	0.638	0.53	0.633	0.635	0.626
Sample size	32	32	32	32	32	32	32	32	32	32	32	32
Variance inflation factor	1.00	4.18	1.02	19.08	20.47	14.73	19.70	15.45	16.65	12.10	17.52	15.95
RESET test	0.116	0.748	0.676	0.158	0.876	0.637	0.719	0.926	0.554	0.075	0.02	0.007

2.9.8 Model selection

We consider it entirely appropriate to triangulate across a diverse suite of models rather than use model selection criteria to remove legitimate models from consideration. As we discuss in Section 4, this approach maximises the information used in the analysis and mitigates the chance that a model unduly benefits/penalises certain companies. However, we do consider that sub-standard models should be eliminated from consideration. For this reason, we apply Ofwat's model selection criteria to eliminate unsuitable models. We implement a points-based system: red is zero points, amber is one point and green is two points. The threshold for a pass is five points. We provide more detail of our assessment in section A.3.

Table 5 - An overview of our model selection process

Model	Plausible sign and magnitude	Statistical validity	Stability/ robustness	Predictive power	Pass/fail
1	Green	Amber	Green	Red	Pass
2	Green	Green	Green	Red	Pass
3	Green	Green	Green	Red	Pass
4	Green	Green	Green	Amber	Pass
5	Green	Green	Green	Amber	Pass
6	Green	Green	Red	Amber	Pass
7	Green	Green	Green	Amber	Fail
8	Green	Green	Red	Amber	Pass
9	Red	Red	Red	Amber	Fail
10	Red	Amber	Red	Amber	Fail
11	Green	Amber	Amber	Amber	Pass
12	Amber	Amber	Amber	Amber	Fail

Models 1-3 do not include an interaction term, meaning they do not sufficiently support the engineering rationale that the interaction between combined sewers and urban rainfall is a key driver of performance. Therefore, while their coefficients have an intuitive sign and they pass the model selection process, we consider that it's appropriate to focus upon models with the interaction term in our final model suite.

Models 6 and 8 fail the robustness test because the scale driver attenuates in the aggregate cost version. Models 9 and 10 fail the plausible sign test because the asset health measure's coefficient is unintuitive. Model 12's density coefficient significantly attenuates leading to an amber score for 'plausible sign and magnitude'. The attenuation is likely due to some correlation between the properties per length driver and FSEs per length.

Therefore, we consider models 4-6, 8 and 11 to be good candidates for triangulation.

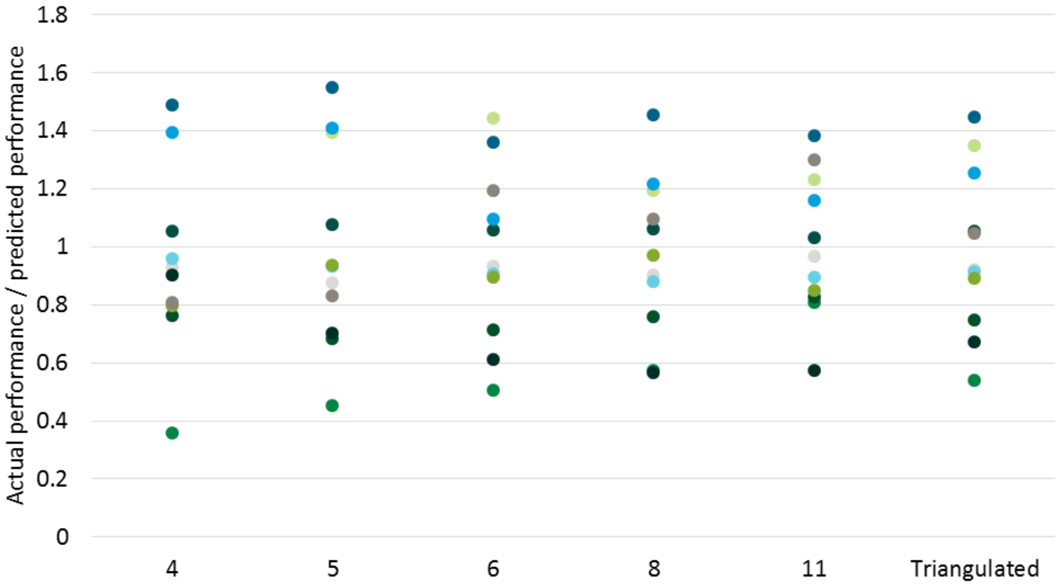
2.10 Set a stretching but achievable performance challenge

Utilising an econometric approach allows us to estimate an upper quartile level of performance, analogous to the way Ofwat sets its catch-up efficiency cost challenge. The upper quartile level of performance would vary to account for the exogenous factors that impact upon performance. We note that the small sample size means some caution should be taken when setting these targets.

When considering which models to use to set performance targets (and indeed cost baselines), we advocate for triangulation across a diverse model suite to capture the different interactions between dependent and independent variables and to mitigate statistical bias in any one model. Figure 9 illustrates the residual spread across the model suite. We note that the residual spread is relatively large when compared to some of the models

Ofwat used during cost assessment at PR19. However, this isn't an unexpected result considering the small sample size and unit performance specification. Figure 9 demonstrates that simple triangulation across our model suite has the effect of clustering companies around predicted performance. Targeted triangulation, which on a company-by-company basis places less emphasis on models with large residual gaps, can reduce the spread further. We discuss targeted triangulated further in Section 4.

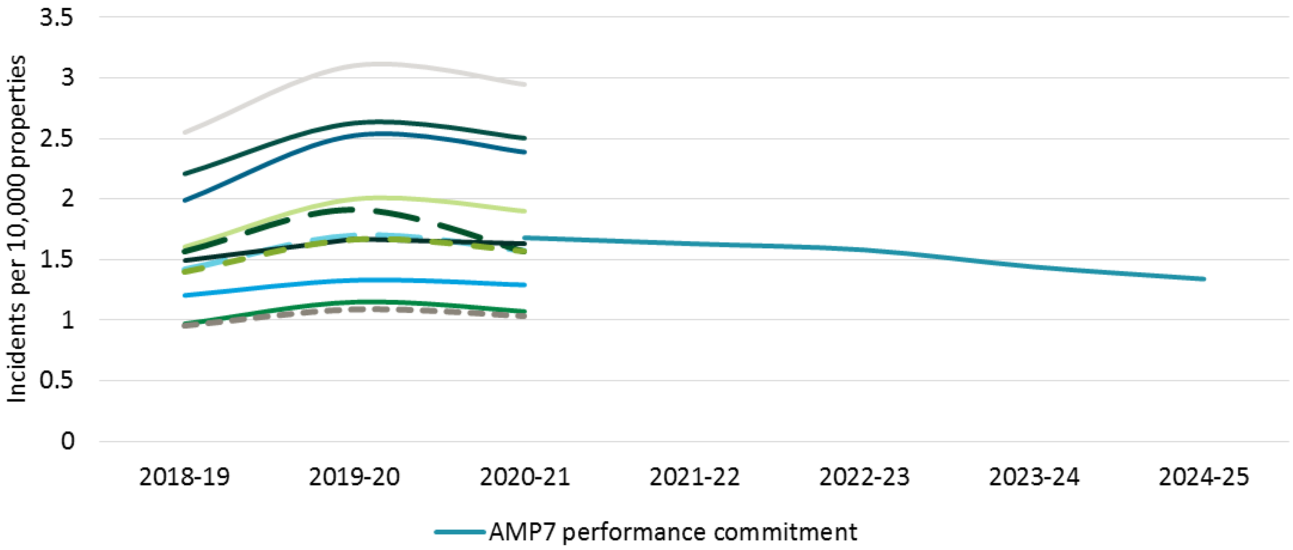
Figure 9 - Triangulation across our model suite clusters companies around predicted performance



We can use the triangulated models to derive performance targets for the industry. We have illustrated this point by estimating the targets at the upper quartile level for transparency, although we note that the small sample size and unit performance specification could justify a less stretching target. Currently, our targets cover the historic period used to develop the models for simplicity. We note that the process could be supplemented with forecasts of the independent variables to derive performance targets in future periods, in the same way that Ofwat's cost baselines develop forecasts of cost drivers.

Figure 10 illustrates the upper quartile level of performance for each company in the industry (triangulated across all 5 models that passed our selection process), and compares this to the AMP7 performance commitment. This upper quartile level is determined by each company's characteristics.

Figure 10 - Upper quartile performance targets compared to the AMP7 performance commitment



It's clear that for some companies, the AMP7 performance commitment is extremely stretching and potentially unachievable, while for others their operating conditions mean that it represents a relatively easy target. This implies that their customers are overpaying for the level of service they receive.

Utilising an econometric approach would allow Ofwat to calculate performance targets that reflect local operating conditions. At PR19, Ofwat assumed that the cost assessment process would be able to account for differences in operating conditions to the extent that a common target was appropriate. However, we are not aware of any analysis that supported this assertion. Our proposed approach would allow Ofwat to compare an equally stretching target for all companies to its common target. This would act as clear evidence to either accept the common target as appropriate or move towards a company-specific target. We consider that this would result in a fairer outcome for both customers and companies.

We also consider that this approach is able to shed light upon the cost-service trade off. For example, it would allow us to assess the scale of investment that would be required to reduce the prevalence of combined sewers and so reduce internal flooding incidents to a particular target. This cost of such investment can then be compared to the service improvement to assess whether it is economic. We also note that the lack of such investment within the historical cost base casts significant doubt on the ability of companies in challenging operating environments to hit the AMP7 target under the PR19 Final Determination.

2.11 A coherent cost assessment approach

We believe this analysis shows there is a strong case for Ofwat to consider a different approach to assessing coherent efficient cost and performance targets where there is evidence that exogenous regional factors drive variations in company performance. Ofwat's framework could address this in two ways:

(1) Adjust base costs to reflect the different challenges faced by companies with different operating circumstances. For example, it could include urban rainfall in its sewer collection and wastewater network plus models. However, this approach could cause companies above the common PC to spend inefficient levels of cost to target unachievable levels of performance and it could result in a much more complex modelling approach.

(2) Adjust performance targets to reflect companies' individual operating circumstances, by implementing the analysis set out above to determine an appropriate performance level for each company. This would keep the cost modelling approach sensibly simple and ensure that all companies receive equally stretching targets.

We consider that option 2 is most appropriate. This would provide a transparent method of setting performance targets while allowing Ofwat to maintain its 'sensibly simple' modelling methodology.

3. Challenging efficiency with a transparent, objective and stable framework

Section 2 explored how an incoherent approach to the cost-service relationship can induce an inequitable productivity challenge on companies across the industry, and suggested a specific solution. This section considers how the PR19 cost assessment framework created implicit productivity challenges more generally.

Regulators mimic the effect of competition by challenging companies to become more efficient. Efficiency assessment presents a number of important issues to consider, and its financial (and emotive) impact means careful thought needs to be given to designing and maintaining a transparent, objective and stable framework if the aim is to produce an outcome that all parties view as legitimate. We note that this is a complex and nuanced area, meaning simple solutions may not be the most appropriate. However, we do consider that a transparent approach should allow companies (and their customers) to explicitly understand the scale of the challenge they are being asked to meet.

3.1 Ofwat's framework was significantly more transparent than at PR14

Overall, Ofwat's framework was significantly more transparent than at PR14. Ofwat engaged with the industry very early through working groups and its methodology consultation. It consulted on its approach to cost modelling and provided detailed commentary explaining its decisions at each stage (IAP, DD, FD). This commentary included the means for companies to replicate Ofwat's assessment, with Stata code and associated Excel models. This was a positive step forwards and should be continued, and where possible built upon, for PR24. We are encouraged by the level of engagement Ofwat is currently driving through its cost assessment working groups and consultations.

One area where we consider Ofwat could improve transparency at PR24 is the experience of the Fast Track companies at the Slow Track DD. Ofwat made significant alterations to its modelling methodology at this stage during PR19, and Fast Track companies were left in the difficult position of being exposed to these changes (as the early certainty principle did not apply to cost benchmarks), while not being explicitly aware of the financial exposure these changes would invoke. Whilst we would support stability within the process as a primary objective, we accept that in some circumstances change cannot be avoided for one reason or another. Therefore, if this were to happen again in the future, we suggest the problem could be mitigated by:

a. giving Fast Track companies cost certainty, similar to the 'do no harm' approach of PR14; or

b. explicitly demonstrating to Fast Track companies the impact of the proposed DD changes at an earlier stage.

3.2 The PR19 framework did not acknowledge implicit sources of stretch

We agree with Ofwat that companies should be pushed to deliver more for customers, and note that Ofwat's framework provided multiple means for Ofwat to induce 'stretch' into the baseline. Some were explicit within Ofwat's framework, but crucially, others were implicit[15]. At PR19, one of the highly contested components in the derivations of botex baselines was the estimate of the rate of frontier shift that Ofwat applied to its model predictions. Ofwat stated[16]:

"There appears to be scope for water companies to improve on-going efficiency…Outperformance of cost baselines during this control period from better performing water companies has been far better than average

[15] We note that we have only considered stretch induced by the cost assessment process, and have ignored stretch imposed by decisions on financing /the WACC.

[16] Ofwat (2019) *Technical appendix 2: securing cost efficiency*. Available here.

performance (better performing companies have outperformed their cost baselines by around 10% since the start of this control period)."

In our business plan submission[17], we explained why an *additional* frontier shift (inclusive of real price effects) of -0.2% (a growth) for Wholesale price controls was in of itself a stretching ask. Some of the key factors were:

- The proposition that historic outperformance against the PR14 baselines justified a more stringent efficiency adjustment. In our view, this logic was flawed as it does not address the fact that outperformance by companies has predominantly been achieved through outperformance of enhancement and not the base expenditure to which the adjustment is then applied.

Figure 11 - UUW analysis of industry AMP6 botex/enhancement performance against the PR14 allowed expenditures

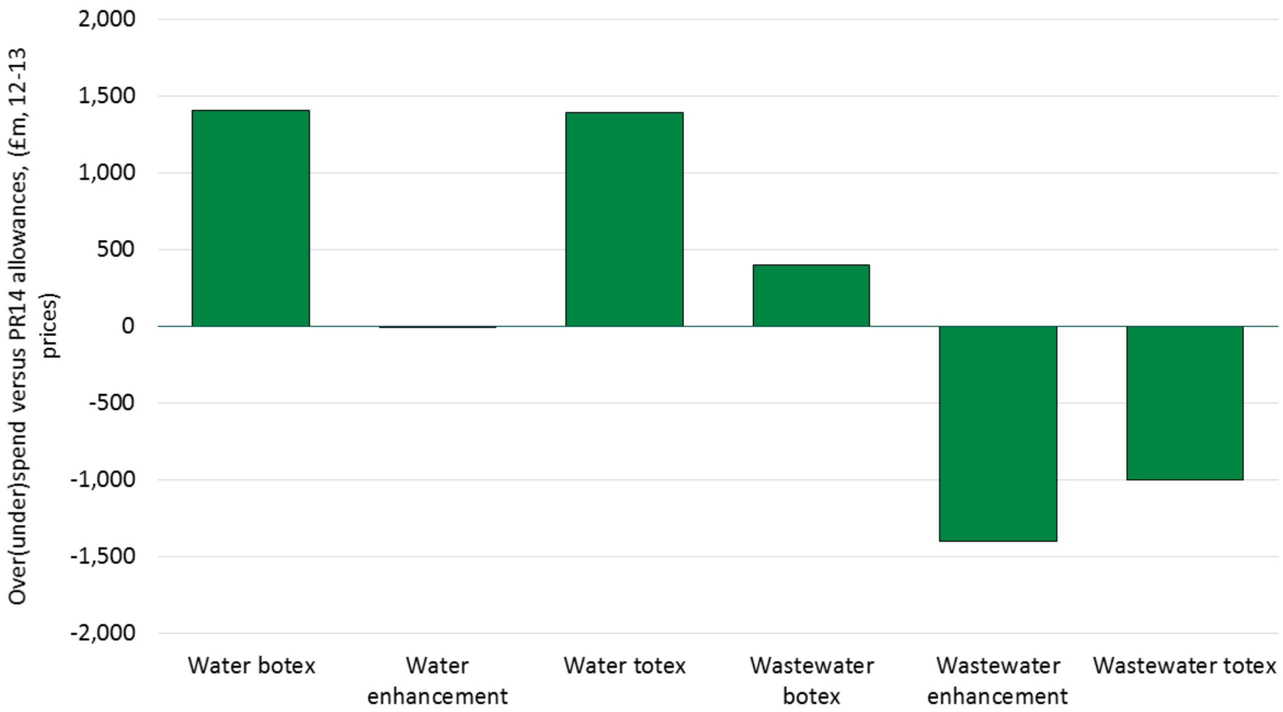


  Figure 11 clearly highlights that whilst some companies may have achieved outperformance within botex, on average we observe base expenditures were expected to outturn higher than that assumed in the PR14 determinations whereas there are significant underspends against enhancement allowances, particularly within Wastewater. We know the issues with modelling enhancement expenditure at PR14, with low quality models leading to over-remuneration for several companies. Additionally, the absence of an industry-wide adjustment mechanism to correct for environmental schemes not subsequently delivered has led to significant outperformance for some companies. While we are unconvinced of the incentive properties of the approach, Ofwat removed the risk of 'overgenerous' enhancement allowances for PR19 through the capping of allowances to be, at most, equal to the business plan, whilst the WINEP cost adjustment mechanism will offer some form of protection against removal of scope within environmental schemes. Therefore, we do not consider that referencing historic outperformance enhancement to justify a frontier shift to base expenditures was appropriate;

- The move from an RPI-linked control to a CPIH-linked control will automatically impart a c1% per annum additional challenge on companies (compared with historic frontier shift assumptions) due to the wedge between the two indices. This transition will naturally cause an increase in real price effects for all companies, as nominal cost fluctuations will not mirror a policy decision by Ofwat;

[17] UUW (2019) *Cost assessment proposal.* Available here.

- It is important to remember that measures of inflation reflect economy-wide productivity gains. Therefore if Ofwat assumed zero (relative to CPIH) for its dynamic efficiency assumption, it would be incorrect to conclude that no dynamic efficiency is assumed – it is just that no additional dynamic efficiency is assumed over and above that being delivered by the economy as a whole;
- It is common for regulators and companies to make use of analyses of Total Factor Productivity (TFP) to inform expectations for dynamic efficiency. Such studies (generally using the ubiquitous EU KLEMS dataset) commonly observe historic trends in productivity gains of between 0% and 1% relative to RPI. KPMG observed similar results in its work for Ofwat at PR19, which indicated a frontier shift of 0.4% to 1.2% for Wholesale. KPMG confirmed its results were evaluated relative to RPI, not relative to CPIH and so rebasing these to be CPIH would lead to a cost growth assumption;
- Pursuing an entirely legitimate modelling methodology of sensibly simplicity while maintaining a high evidential bar for cost adjustment claims reduces the likelihood that companies' benchmarks reflect the challenges their operating regions pose, effectively constituting a productivity challenge; and
- Whilst it is obviously an attractive proposition for customers, Ofwat's requirement that companies must achieve the delivery of more stretching performance through base costs (for example, the 15% leakage reduction[18]) means that companies were already implicitly accepting a significant frontier shift without a further adjustment to the baseline. The actual frontier shift required will differ from company to company depending on their changes in performance and regional operating conditions but it is significantly greater than zero and a coherent framework should not overlook this when implementing the forward-looking challenge.

The additive effect of each of these challenges not accounted for within cost assessment means that the underlying productivity a company must achieve in order to meet Ofwat's baseline view of base expenditure is not 1.1% per annum but is significantly higher. We estimate that the real value of productivity could comfortably be greater than 6.5% per annum given the:

- ~1% per annum challenge due to the switch to CPIH;
- ~1% per annum challenge as a result of the exclusion for any real price effects within power and chemical costs;
- >3% per annum challenge for achieving enhanced levels of service through base service allowances; and
- 1.1% per annum challenge through Ofwat's explicit frontier shift.

All companies were asked achieve these record productivity improvements on base expenditures on day one of the AMP (there was no glide path), even though history indicates that simply achieving a (static) upper-quartile has proven difficult for many. It is enhancement expenditure where companies can drive out the biggest efficiencies, potentially due to the abilities to make better use of markets to deliver solutions and so placing additional challenge on base expenditures may be counterfactual. Overall, we consider there is considerable scope for the cost assessment framework to better acknowledge the various sources of stretch at PR24, and as we stated in section 2, this should also recognise that regional operating circumstances can lead to any implicit sources of stretch being inequitably distributed across the industry.

[18] Wholesale Water Supply-demand balance enhancement – feeder model summary p12, Ofwat

4. A diverse model suite better reflects the costs of delivering water and wastewater services

Utility network providers can provide multiple services through a single value chain. Having a clear definition and understanding of the services provided enables us to explicitly target cost assessment towards a legitimate outcome.

Different companies will structure their operations differently to optimise based on local operating requirements when delivering the same service. This means that efficient spend within each service can vary across the industry, driven by factors that are outside of management control. Failure to account for these differences within the cost assessment framework could result in efficient companies being penalised or customers of inefficient companies (unduly) paying more. The ability of the cost assessment framework to reflect these complex underlying cost drivers is determined by the approach taken to the benchmarking model suite and cost adjustments. A diverse model with a varied basket of cost drivers is more able to reflect the costs of providing each service to customers.

At PR19, Ofwat pursued sensible simplicity in its model suite. This meant it focused upon a smaller number of models that used a limited pool of variables to reflect the four main cost drivers identified by Ofwat: scale, complexity, density and topography. This approach represented a considerable improvement to that taken at PR14, where Ofwat implemented more complex transcendental logarithmic (translog) models. Considered in isolation, Ofwat's PR19 model suite was reasonable and received support from the wider industry and the CMA.

However, while a 'sensibly simple' modelling approach can be legitimate, the consequence is that it should be expected that companies would need to submit a larger number of cost adjustment representations to ensure that their individual circumstances are properly reflected in the 'sensibly simple' benchmark. This should not be interpreted as companies seeking inefficient funding but rather as an inexorable corollary of an entirely legitimate methodological choice of parsimony.

A coherent framework should recognise this relationship between the simplicity inherent in the modelling approach and the subsequent inevitable need for cost adjustments if the aim is to ensure companies are allocated efficient costs commensurate to the operational challenges they face. Alternatively, the regulator can pursue a policy of diversity in the model suite. This approach slightly increases the complexity of the model suite relative to a 'sensibly simple' approach. However, it does reduce the need for cost adjustment claims.

4.1 The benefits of a more diverse model suite

It is impractical to expect that any one model or any suite of models is capable of perfectly capturing the relationship between cost and cost driver across the industry with the data available. The heterogeneity of operating conditions across the water industry means that a more diverse approach is much more capable of effectively recognising efficient costs for each company. This is because:

- It maximises the information that is used to build the benchmark, which should increase the engineering, operational and economic rationale underpinning the cost assessment framework;
- There is more potential to capture key cost drivers across the different services that companies provide through the same value chain, or capture the cost driver in an alternative way using a different variable. This should reduce the risk of the choice of variable unduly benefiting/penalising certain companies;
- The need for cost adjustments decreases. This places more weight on the modelled benchmark, which is a well-understood and transparent assessment method. This should help to reduce problems relating to information asymmetries between regulator and companies. It also means that both regulator and companies realise substantial time and cost savings related to writing and assessing cost adjustment claims; and
- It allows different aggregations of the value chain to be included, which will provide a better understanding of substitution effects across the value chain within the benchmark.

4.2 How did Ofwat approach diversity at PR19?

While Ofwat's approach to modelling at PR19 was a big improvement to the approach taken at PR14, we consider that there is considerable scope for the PR24 framework to adopt an enhanced approach in recognising the diverse operating conditions across the industry.

At PR19, Ofwat made a policy decision to pursue a 'sensibly simple' model suite. It identified four key cost drivers and sought to reflect these in its models using a parsimonious selection of cost drivers. Sensible simplicity is an entirely legitimate methodology but it does increase the likelihood that companies with particularly adverse operating conditions would be significantly penalised, particularly given the level of stretch Ofwat imposed elsewhere in its framework. A coherent framework should recognise the relationship between the modelling methodology and the subsequent impact upon the cost adjustment assessment process.

We do consider that Ofwat adopted an appropriately diverse set of cost aggregations. Ofwat utilised an aggregation called 'water resources plus', which comprised water resources, raw water distribution and water treatment. This aggregation recognised the substitution effects that exist across these different services; companies with comparatively cheaper water resources (gravity fed impounding reservoirs and river abstractions) tend to face higher relative treatment costs due to poorer raw water quality. Companies with comparatively more expensive water resources (e.g. groundwater that requires pumping from boreholes) will face lower relative treatment costs because the water quality from their sources is better and so typically requires less complex treatment. Modelling the costs from these value chains together meant that the benchmarking models reflected this trade-off as the substitution effects net off from one another.

We also supported Ofwat's use of the 'Bioresources Plus' modelling split at PR19 because this captures the substitution effects that exist between Wastewater Network Plus and Bioresources due to differences in asset configuration across the industry. For this reason, we do not support Ofwat's proposal to model Bioresources completely separately from wastewater treatment at PR24[19]. This will mean that the benchmark is inappropriately impacted by company decisions on asset configuration rather than solely by efficiency.

At PR19, Ofwat also accommodated the existence of substitution effects across the different value chain elements through its approach to calculating the catch-up efficiency challenge. It calculated the catch-up efficiency challenge at the aggregated level (either water or wastewater). If Ofwat had calculated the efficiency challenge at a more disaggregated level (for example, by setting separate and specific Bioresources and Wastewater Network Plus efficiency challenges) then its benchmarks would not have recognised that the presence of substitution effects mean that the lowest cost company in one area may actually be a high cost company in another[20]. The effect of this would be that the combined benchmark would be unobtainable by any one company as the challenges would also be due to regional operating circumstances and/or asset locations rather than differences in efficiency. As a result, we support assessing the efficiency challenge at the aggregate level as adopted by Ofwat at PR19.

4.3 How to increase diversity within the benchmarking process

Diversity can be achieved through two different approaches. We advocate that both approaches be implemented simultaneously:

- We can utilise a diverse set of explanatory factors. This could allow us to consider additional cost drivers within the model suite, or reflect existing cost drivers using alternative variables. This should mitigate the risk that certain cost drivers favour certain companies and penalise others. This approach would increase the number of models used to explain each aggregation of cost, with the outcome being a benchmark that draws upon a richer variety of information.

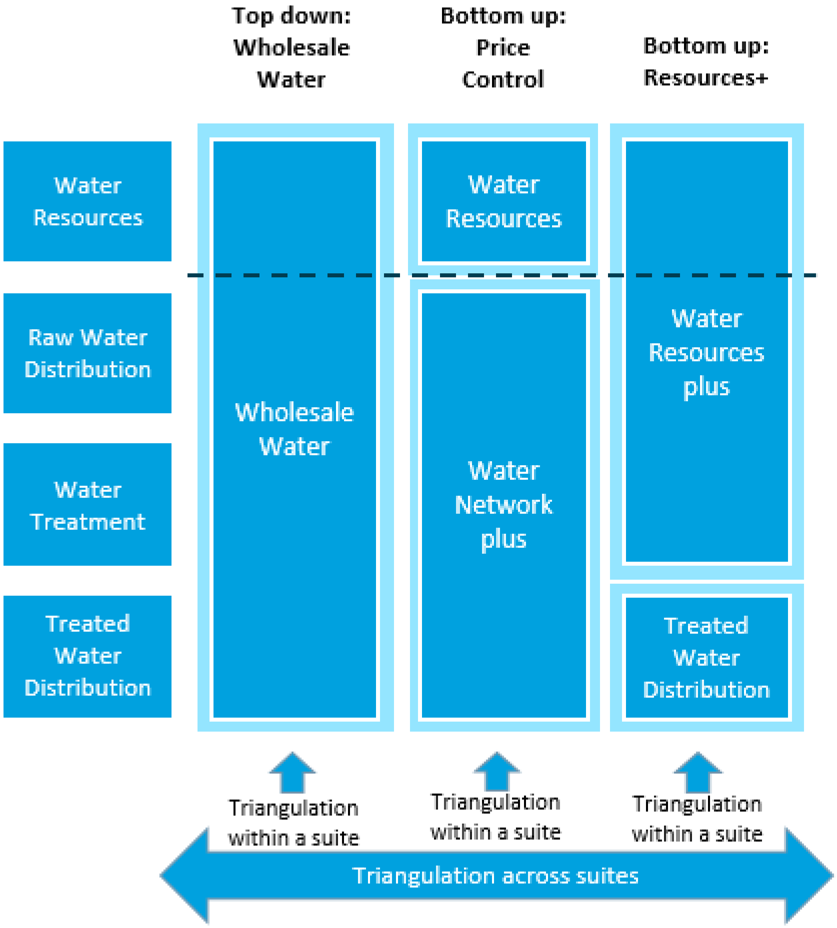
[19] Ofwat (2021) *Assessing base costs at PR24.* Available here.

[20] The benchmark would also be influenced by differences in cost allocation practices, which is, again, undesirable.

- We can model different aggregations of cost. This will help us develop a better understanding of the relationship between cost and cost driver across different parts of the value chain. It will also reflect the presence of 'substitution effects' in the benchmark.

In our cost assessment proposal at PR19[21], we utilised a diverse suite of models with a wide range of explanatory variables which sought to capture the complex relationships between cost and cost drivers across the whole value chain. This approach is illustrated in Figure 12. We would support a similar approach at PR19.

Figure 12 – An example of triangulation within and across model suites



4.4 Using targeted triangulation to capture diversity when triangulating between models and model suites

A more diverse model suite raises questions on best to incorporate the different predictions. Simple averaging implicitly assumes that all models and all suites are equally suited to predicting expenditure and that no one approach can generate a more realistic baseline for a company. The greater the degree of homogeneity between the companies then the more appropriate this assumption becomes as the resulting model predictions should account for all companies circumstances equally. However, if company circumstances are more heterogeneous, or if there are threats to the internal validity of the models caused by omitted variable bias or errors, then triangulation on an individual company basis can be a useful tool to use to minimise the risks of extreme outliers as a result of model deficiencies.

There are a number of instances where regulators have historically used a 'targeted' approach to triangulation as a method of correcting for modelling deficiencies. The German energy regulator, Bundesnetzagentur, utilises both

[21] UUW (2018) *S6002: Cost assessment proposal*. Available here.

SFA and DEA[22] approaches when developing their benchmarks for each distribution system operator (DSO). Accepting that neither statistical method is perfect and that the different approaches will have varying success in predicting required expenditure for individual companies, it determines the actual efficiency score (ES) using a 'best off approach' whereby the maximum efficiency score is applied, with a minimum of 60%.

$$ES = \max(DEA_1, DEA_2, SFA_1, SFA_2, 0.6)$$

The same principal occurs within Austrian gas distribution regulation but in this case, the weighting assigned to the 'best off' estimation method (DEA and MOLS[23]) is pre-determined using a 60/40 ratio with the larger weighting applied to the models that give the highest ES.

$$ES = 0.6 \times \max(DEA, MOLS) + 0.4 \times \min(DEA, MOLS)$$

Whilst we recognise that such weightings are somewhat arbitrary, we agree that it is unattainable to expect that any one model can act as a perfect prediction for every company within the industry. As a result, an appropriately triangulated diverse suite of models is the most suitable method to pursue in order to ensure that the resulting baselines are a fair reflection of each company's required expenditure. Rather than arbitrarily assigning a greater weighting to the suite/model that result in the *largest* prediction, we believe that minimising the residuals (and therefore equally addressing both over and under predictions) is the most appropriate approach to take as it prevents baselines being set too high because of a model/suite overestimating the requirements.

To achieve this, we developed the following matrix whereby weightings (Figure 13 showing a 50% gap between the 'best' and 'worst model/suite and Figure 14 showing 100% gap), whether they are applied at a model level or a suite level, scale linearly based on the performance and minimising the resulting residuals for each company. The matrix determines the 'gap' between the weightings of the 'best' and 'worst' model or suite. The resulting weights (in the blue box) can then be applied to the suites or models based on the rankings of the residuals for each company (the ratio between actual cost and modelled cost).

Figure 13 - UUW proposed triangulation matrix using 50% gap between high and low weightings

Number of models/suites to trianguate	2	3	4	5	6
Gap from high to low	50%	50%	50%	50%	50%
Start (pre normalisation)	*100%*	*100%*	*100%*	*100%*	*100%*
	150%	*125%*	*117%*	*113%*	*110%*
		150%	*133%*	*125%*	*120%*
			150%	*138%*	*130%*
				150%	*140%*
					150%
Total (for normalisation to 1)	250%	375%	500%	625%	750%
Normalised weightings	*40%*	*27%*	*20%*	*16%*	*13%*
	60%	*33%*	*23%*	*18%*	*15%*
		40%	*27%*	*20%*	*16%*
			30%	*22%*	*17%*
				24%	*19%*
					20%
Check	*100.0%*	*100.0%*	*100.0%*	*100.0%*	*100.0%*

[22] Stochastic frontier analysis (SFA) and data envelopment analysis (DEA) are alternative statistical approaches for the estimation of comparative benchmarks.

[23] MOLS = modified ordinary least squares

Figure 14 - UUW proposed triangulation matrix using 100% gap between high and low weightings

Number of models/suites to trianguate	*2*	*3*	*4*	*5*	*6*
Gap from high to low	100%	100%	100%	100%	100%
Start (pre normalisation)	*100%*	*100%*	*100%*	*100%*	*100%*
	200%	*150%*	*133%*	*125%*	*120%*
		200%	*167%*	*150%*	*140%*
			200%	*175%*	*160%*
				200%	*180%*
					200%
Total (for normalisation to 1)	300%	450%	600%	750%	900%
Normalised weightings	*33%*	*22%*	*17%*	*13%*	*11%*
	67%	*33%*	*22%*	*17%*	*13%*
		44%	*28%*	*20%*	*16%*
			33%	*23%*	*18%*
				27%	*20%*
					22%
Check	*100.0%*	*100.0%*	*100.0%*	*100.0%*	*100.0%*

We have illustrated the results of 50% and 100% weightings but the scaling can be set to any level which best reflects the suitability of the models/suites to predict costs for all companies. A higher weighting places a larger differential between the ‘best’ model/suite and the ‘worst’ model/suite that will minimise the sum of the residuals further. Economic priors and the spread and magnitude of the residuals will largely inform the most appropriate final weighting requirements.

The benefit of adopting this approach over simple averaging or arbitrarily weighting based on resulting predictions is that averaging implies that each model/suite is equally adept at predicting requirements for all companies (which we know not to be true) and an arbitrary weighting will simply result in predictions increasing for all companies which is to the detriment of customers. In contrast, weighting models/suites by seeking to minimise the residuals will prevent models or suites that are too generous for a company obtaining as high of a weighting whilst minimising undue challenge due to a particularly poor performing model/suite.

Appendix A More details on our approach to modelling internal sewer flooding performance

A.1 Our independent variables

Our assessment of engineering, operational and economic performance drivers created a pool of potential exogenous variables to test within econometric models. This section sets out our proposed specification of these variables.

Combined sewers

- This is calculated by dividing 'Length of formerly private sewers and lateral drains (s105A sewers)' (BON ref BN13528) by the sum of 'Total length of "legacy" public sewers as at 31 March' (BON ref BN13535) and 'Length of formerly private sewers and lateral drains (s105A sewers)' (BON ref BN13528), multiplied by 100.
- We note that South West revised its figures for formerly private sewers and lateral drains in APR21. This has influenced its values for other variables that are functions of total sewer length.

Urban run-off

- We use the urban rainfall variable that was developed by Arup and Vivid Economics at PR19, normalised by sewer length.

Urban run-off x combined sewers

- We multiply urban rainfall by the percentage of combined sewers to create the interaction term. The addition of the interaction term means that the impact of urban rainfall on internal sewer flooding now depends upon the level of combined sewers and vice versa.
- Note that we do not use normalised urban rainfall to create the interaction term. This is because sewer length would appear in the denominator of both terms (combined sewers/sewer length, urban rainfall/sewer length) meaning that multiplying the two together would create a squared denominator. Therefore, we use total urban rainfall to create the interaction term.
- A positive interaction term would imply that a more extensive combined sewer network causes urban rainfall to have a greater impact upon internal flooding incidents (and vice versa). This would align with engineering, operational and economic rationale.

Weighted average density

- We use Ofwat's weighted average density measure from PR19. This calculates the weighted average number of people in each square kilometre of a company's area.
- We have extended the dataset to include the additional years since PR19.

Properties per kilometre

- We use Ofwat's alternate measure of density from PR19, calculated as total number of properties (in units of 10,000) divided by total sewer length. For property numbers, we used BON ref BN1178 pre-APR21 and BON ref BN2710 from APR21 onwards.

Number of pumping stations per kilometre

- We divide total number of pumping stations (BON ref S6019) by total sewer length to capture the effect of topography and sewage maceration.
- We prefer the number of pumping stations to capacity of pumping stations, because a hilly area could require a larger number of smaller pumping stations rather than one very large one (although this depends upon the specific characteristics of a network so admittedly may not hold in all cases). This would then have the effect of macerating sewage across a larger area of sewer.

Sewer collapses

- We divide total sewer collapses (BON ref 13521) by total sewer length to capture the effect of asset health. This is because a sewer length with more collapses per km indicates poorer asset health.

Food service establishments per kilometre

- We use government data on Food Service Establishments (FSEs) by local authority. This sample was collected in 2018 and has not been refreshed since. This means we have to assume that the number of FSEs is unchanged in every local authority for each year of our sample.
- We aggregate FSEs by local authority to a company level using the mapping that Ofwat used to derive its density driver at PR19. We normalise by sewer length because additional FSEs per length of sewer can be expected to increase flooding risk.

A.2 The interaction between combined sewers and urban rainfall

The introduction of an interaction term changes the way we interpret the coefficients in the model. The impact of urban rainfall upon internal sewer flooding now depends upon the level of combined sewers and vice versa. To see this, consider the following equation:

$$Internal\ floods = \beta_0 + \beta_1.combined\ sewers + \beta_2.rainfall + \beta_3.(combined\ sewers \times rainfall) + \varepsilon$$

A marginal change in rainfall has the following effect on internal sewer flooding (by the rules of differentiation):

$$\frac{\partial(internal\ floods)}{\partial(rainfall)} = \beta_2 + \beta_3.\text{combined sewers}$$

This makes it clear that the effect of urban rainfall on internal floods depends upon the level of combined sewers. For comparison, a model without an interaction term has a different interpretation:

$$Internal\ floods = \beta_0 + \beta_1.combined\ sewers + \beta_2.rainfall + \varepsilon$$

Under this model structure, rainfall has the following impact on internal sewer flooding:

$$\frac{\partial(internal\ floods)}{\partial(rainfall)} = \beta_2$$

Which is more aligned to the interpretation of Ofwat's model suite at PR19, where coefficients could be interpreted in isolation. By contrast, where models include an interaction term, it can be misleading to try to interpret coefficients in isolation. This is because such an interpretation requires the other term in the interaction term to equal zero. To see this, consider the following model:

$$Internal\ floods = \beta_0 + \beta_1.combined\ sewers + \beta_2.rainfall + \beta_3.(combined\ sewers \times rainfall) + \varepsilon$$

In order for us to interpret β_1 independently, we would have to assume rainfall is zero:

$$Internal\ floods = \beta_0 + \beta_1.combined\ sewers + \beta_2.0 + \beta_3.(combined\ sewers \times 0) + \varepsilon$$

$$\rightarrow Internal\ floods = \beta_0 + \beta_1.combined\ sewers + \beta_2.0 + \beta_3.0 + \varepsilon$$

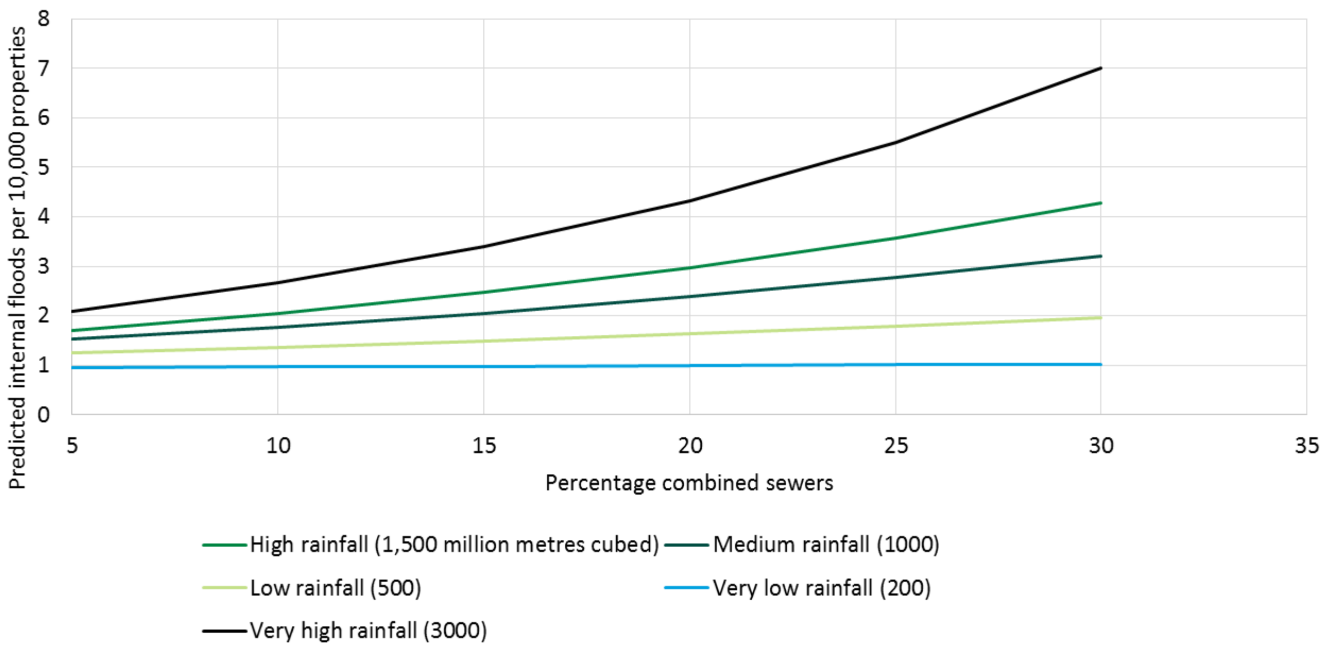
$$\rightarrow Internal\ floods = \beta_0 + \beta_1.combined\ sewers + \varepsilon$$

$$\rightarrow \frac{\partial(internal\ floods)}{\partial(combined\ sewers)} = \beta_1$$

However, assuming zero annual rainfall across an entire company's area is clearly unrealistic. Therefore, care needs to be taken when interpreting models with an interaction term.

The easiest way to understand what effect the interaction term has is through a graph. Figure 15 demonstrates how the effect of combined sewers on internal floods varies as the amount of rainfall experienced varies using the coefficients from model 4. It's clear that combined sewers has a bigger impact upon internal sewer flooding when there is more urban rainfall.

Figure 15 - Interpreting the effect of the interaction term



A.3 Our model selection process

Ensure the estimated coefficients are of the right sign and of plausible magnitude

- Models where coefficients are in line with our engineering, operational and economic rationale receive a green score.
- Models where any coefficient is unintuitive receive an automatic fail. This ensures that unintuitive coefficients do not skew the analysis. Note that coefficients on variables within the interaction term were assessed as a whole, not individually.

Evaluate the statistical validity of the model across a range of statistical diagnostic tests

- We assess the statistical significance of each coefficient, and consider the Ramsey RESET test and the average Variance Inflation Factor. We are more tolerant of a high VIF score in models with an interaction term, in line with Ofwat's approach to models with a quadratic term in at PR19.
- As we are using a small dataset, we are willing to tolerate a lack of statistical significance at conventional levels as long as the coefficients have an intuitive sign. This is in line with Ofwat's approach at PR19.

Assess if the estimated model results are stable / robust to changes in the underlying assumptions and data

- One approach to robustness testing is to systematically drop observations to test whether individual observations have an outsized impact on the estimates. However, we only have 32 data points in our sample. This means that a robustness test that drops individual observations may not be telling us much about model robustness, rather it may just be reflecting the effect of sample size. For this reason, we do not implement such tests.
- Therefore, our primary forms of robustness testing are to change the model specification from a unit cost model to an aggregate cost model (where scale is reflected as an independent variable rather than as the denominator of the dependent variable) and to test other estimators. Models that retain intuitive coefficients and do not lose statistical significance pass this test.

Assess the predictive power of the model

- We note that there are some limitations to assessing predictive power in a model that covers a relatively short time-frame. However, we note that our focus upon engineering, operational and economic narratives should mean that our models have reasonable predictive power, even in a small sample. This would especially be the case if we triangulated across a diverse suite.

- Relying on R squared values could be misleading. This is because this measure will always increase as more independent variables are added, meaning that a selection process that solely relies on it may lead us to favour over-fitted models.
- For these reasons, we score all models an amber unless there is a compelling reason for an alternative score. As more data is added in future years, we may be able to refine the way this test is applied in this context.

A.4 More detail on the sewer flooding 'hackathon'

A hackathon is a design sprint event in which developers, interface designers, project managers, domain experts, and others collaborate intensively on projects.

The purpose of our sewer flooding hackathon was to determine whether a hackathon approach could be used to investigate flooding risk using new and previously utilised datasets together with the application of data science methods. We hoped to develop a better understanding of the root cause of incidents, future risks and the effectiveness of interventions.

The hackathon identified 47 insights into sewer flooding performance. We have set out five key insights below:

- Food service establishments increase flooding other causes (FOCs) risk.
- Surface water interactions increase flooding risk in particular areas.
- Cellar density acts to increase FOC risk.
- Combined sewers act to increase flood risk.
- Seasonality and rainfall impact FOC risk.

The findings of the hackathon strongly influenced our assessment of the engineering, operational and economic rationale for internal sewer flooding performance drivers.

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

unitedutilities.com

United Utilities

Water for the North West