Future Ideas Lab

The Principles of Regulatory Cost Assessment

July 2021

A submission to Ofwat's Future Ideas Lab setting out the principles of regulatory cost assessment

United Utilities

Water for the North West

Executive Summary

A key part of the periodic price review involves the regulator setting the efficient cost of service delivery. In doing so, it makes an assessment as to whether the company's proposed costs are efficient or whether intervention is needed in order to protect customers. This process is known as cost assessment.

The purpose of this paper is not to attempt to establish the 'best' technical approach to delivering cost assessment; indeed there is unlikely to be a 'one-size-fits-all' approach. Instead, we believe that there is better value in first establishing how an objective framework, grounded in sensible principles, should be developed. This is because cost assessment can draw upon a variety of powerful and flexible tools, but it does involve subjective decisions and trade-offs in how best to employ them. It is important to ground these decisions in a set of objective principles to ensure a legitimate outcome for customers and companies. This is what we seek to establish in this paper.

This paper forms our first cost assessment contribution to the Future Ideas Lab. We intend to publish a series of contributions relating to cost assessment, which we summarise below:

1. **Paper 1** – we draw upon our experience of PR19 to define six principles that we consider will lead to a legitimate outcome during any cost assessment process.
2. **Paper 2** – we examine Ofwat's approach at PR19 against the principles set out in Paper 1, and the implications of this approach at PR24.
3. **Paper 3** – we consider what changes are required to increase the external validity of the models in response to possible changes in Ofwat's approach/focus e.g. natural capital, zero net carbon, partnership working.
4. **Paper 4** – we draw upon the previous papers to make more comprehensive proposals for PR24 cost assessment.

We have identified six key principles of cost assessment. These principles have been shaped by our reflection on cost assessment at PR19, and what potential challenges lie ahead at PR24. We were generally supportive of Ofwat's approach to cost assessment at PR19 and consider it a significant improvement to PR14. We seek to build upon these foundations.

Principle 1: Define the services provided

Utility network providers can provide multiple services through a single value chain and this should be explicitly accounted for within the cost assessment framework. Only once service definition is fully understood can the regulator begin to develop the appropriate tools to estimate the cost of providing each service. Having a clear definition and understanding of the services provided enables us to explicitly target cost assessment towards a legitimate outcome.

Importantly, 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 in question, 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.

Principle 2: Prioritise engineering and economic rationale within cost assessment

The process of deciding upon appropriate cost drivers to use within cost assessment should be built upon a deep understanding of the engineering and economic narratives underlying the defined service provided. The intuition behind this derives from the statistical concepts of causation and correlation. Correlation describes the size of relationship between two different variables, whereas causation reveals that one variable is directly affected by the other.

To facilitate this, there should be a clear *a priori* relationship between the chosen set of cost drivers and observed cost. This should be underpinned by objective model selection criteria, which prioritise engineering and economic rationale.

Principle 3: Protect the benchmark's independence

The competitive market that regulators seek to emulate is characterised by participants being unable to influence the outcome, in terms of the market price or the costs of their competitors. This means that for regulators to replicate the incentives present in a competitive market, regulated companies should not be able to influence the outcome of cost assessment through management decisions on how to structure or operate the company. This also ensures that the regulator's benchmark is wholly independent.

Independence is essential because it means that the benchmark is not influenced by the company's plan. This ensures that there are the appropriate incentives for companies to reveal efficient costs up front, whilst promoting the best outcome for customers.

Principle 4: Ensure expenditure outside of the modelled historical period.

Cost assessment should recognise all services within a company's past and future operations and key operating and environmental differences. This means that it cannot be a solely backwards looking assessment based on what has been delivered previously; it also needs to account for future changes to the market. The most effective way of doing this is through emphasising external validity, which reflects how well cost assessment can be extrapolated into future periods.

In contrast, internal validity focuses on the statistical fit of the model. While this is an entirely valid approach, we consider that external validity should be prioritised within cost assessment as this maximises the likelihood that the regulator's benchmark is able to appropriately reflect cost drivers throughout the next price control period. Internal validity should be used as a sense check to choose the best performing 'externally valid' model.

Principle 5: Ensure there is a coherent approach to cost benchmarking and the wider framework

A key challenge faced by the regulator is ensuring that its cost assessment methodology is joined-up; it must be mindful that decisions in one area can have consequential impacts in another area. This is particularly true when considering how the cost adjustment and enhancement framework best complements the benchmarking approach.

We consider that the approach to cost adjustments should be driven by the approach taken during the benchmarking process. If benchmarking draws upon a wide variety of cost drivers and information, such that the expectation is that different company circumstances are reasonably represented within the benchmark, then the need for cost adjustments should be reduced (presuming the benchmark has external validity).

If, however, a more parsimonious or 'sensibly simple' approach to developing the benchmark is taken, then the consequences are that companies can be expected to submit a larger number of cost adjustment representations to ensure that their individual circumstances are reflected in the benchmark. The regulator should not interpret this as companies seeking inefficient funding but rather as an inexorable corollary of its entirely legitimate methodological choice of parsimony.

We also consider that a coherent framework necessitates an appropriate distinction between base and enhancement expenditure. The botex models used at PR19 relied on cost drivers that were relatively static over time. While such cost drivers are largely able to predict base expenditure appropriately, enhancement expenditure is much more 'lumpy', meaning botex models struggle to appropriately reflect such expenditure in the benchmark. Compounding this problem is the top-down nature of botex models; an overall allowance is provided, meaning neither company nor regulator knows exactly how much has been funded for a particular activity. For this reason, we advocate that base and enhancement costs should generally be assessed separately.

Principle 6: Challenge efficiency with a transparent, objective and stable framework

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 an objective framework if the aim is to produce an outcome that all parties view as legitimate.

Methodological choices made in cost assessment can influence the regulator's assessment of efficiency gains attainable by companies. Therefore, an objective framework for assessing efficiency should not seek to make ex post alterations in response to a perceived level of stretch that is different to prior expectations. Rather, the regulator should be fully cognisant of how legitimate methodological choices can lead (entirely reasonably) to a reduced catch-up challenge. Additionally, the regulator should acknowledge the overall stretch (explicitly and implicitly) being applied by the cost assessment framework to ensure that its benchmark is realistic.

Contents

What is Regulatory Cost Assessment?

Utilities such as water, electricity and broadband tend to be provided by a single regional or national company. This is because the significant cost involved in setting up mean it would be inefficient for another company to replicate the asset base in order to compete with the established providers. Such industries are known as natural monopolies.

THE PRICE REVIEW PROCESS

1. Set stretching service levels
2. **Establish efficient cost of delivery**
3. Work out efficient financing costs

These natural monopolies therefore tend to be subject to independent regulation, which ensures customers receive a good service at a fair price, by mimicking the effect of competition. The Price Review is central to this process. During the review, the regulator sets the price and performance package that customers receive for the next regulatory period.

THE PROFIT INCENTIVE

- In a competitive market, companies are incentivised to minimise costs and improve service in order to gain market share and maximise profit.
- Comparative competition and associated incentives needs to be introduced in regulated 'natural monopoly' industries.

A key part of the price review involves the regulator setting an assumed "benchmark" level of cost. In doing so, it makes an assessment as to whether the companies' proposed costs are efficient or whether intervention is needed in order to protect customers. This is a key objective of the Price Review; to ensure that customers pay an appropriate price and receive a good quality service.

Cost assessment is more than just about the development of econometric models or making adjustments for future efficiency (although these topics often receive the most airtime); it is the overarching framework through which the regulator is able to determine the cost that customers will ultimately pay. It will likely involve many different complex estimation methods and supplementary processes and as such, the regulator must ensure that there is alignment between all of its different approaches to ensure that the result is fit for purpose.

"The aim of economic regulation is to create a system of incentives and penalties that aim to replicate the outcomes of competition in terms of consumer prices, quality and investment."
Department for Business, Innovation and Skills (2010) *Principles for Economic Regulation*

The importance of independence

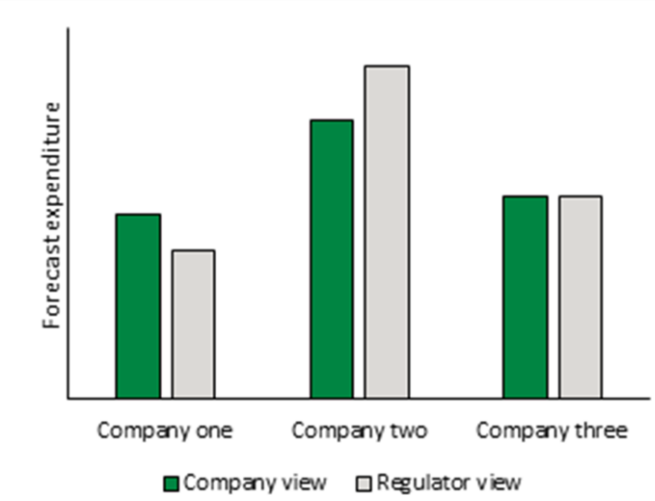
A regulated company has better information about its cost base than the regulator - this is called information asymmetry. Absent any correction for this problem, the regulated company might not have as strong an incentive to reveal its efficient costs within its business plan, thereby risking customers paying too much.

To avoid this problem, regulators often establish an **independent benchmark** to assess the efficiency of a company's planned expenditure. Independence is essential because it means that the companies are less able to influence its value. This ensures that there are the appropriate incentives for companies to reveal efficient costs up front, whilst promoting the best outcome for customers. This aligns with the idea that any company in a competitive market needs to set its prices by reference to the market price or the costs of its more efficient competitors.

Benchmarking is a tool that regulators across the world use to simulate the effect of competition, whereby company targets are set by reference to better performing comparator companies. This replicates what happens in a market, where in order to remain competitive, a poorer performing company may need to set lower prices informed by a better performing competitor, despite having higher costs.

It is important to realise that independent benchmarking can lead to situations where a company's proposed costs are lower than the regulator's view. This does not serve to undermine the legitimacy of the benchmark. Rather, it may simply demonstrate that a company has proposed a more efficient business plan, which is an important feature of an incentive-based system of regulation i.e. that better performing companies are able to earn additional profit.

REGULATORY BENCHMARKING



This chart illustrates the range of outcomes possible in regulatory benchmarking, where:

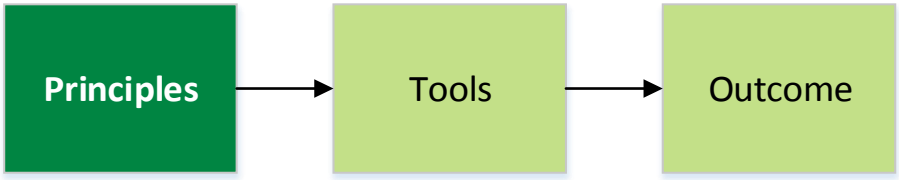
- Company one has submitted an inefficient plan, and will have its costs challenged.
- Company two has submitted an efficient business plan, and revealed its efficient level of cost to the regulator. It continues to receive the higher allowance

The overall effect of a regime like this is to incentivise companies to reveal efficient costs, which ultimately passes the benefit on to customers over time.

Regulatory benchmarking can involve the use of statistical models to make comparisons between companies, which are used to estimate the efficient level of expenditure for an individual company. In theory, these models abstract away from management decisions meaning they produce estimates of efficient spend only, which supports the development of an independent benchmark. Whilst this paper primarily relates to the use of cost models, there are other tools that regulators can draw upon when benchmarking, such as a bottom-up engineering assessment.

Cost assessment can draw upon a variety of powerful and flexible tools, but it does involve subjective decisions and trade-offs. Therefore, it is important to ground these subjective decisions in a set of objective principles to ensure a legitimate outcome for customers and companies.

Figure 1 The context of setting out principles within the wider cost assessment framework



Objectives of this paper

The purpose of this paper is not to attempt to establish the 'best' technical approach to deriving an independent benchmark, indeed there is unlikely to be a 'one-size-fits-all' approach. Instead, we believe that there is better value in first establishing how an objective framework, grounded in sensible principles, should be derived. Where we note one potential methodological avenue, we have sought to identify equally legitimate alternative choices and clarify what factors might influence the choice of one over the other and what consequent decisions need to be made. In this way, we intend this framework to be able to adapt to future regulatory challenges and technical approaches.

Without clear guiding principles, any un-contextualised subjectivity within a regulatory decision might lead to an outcome considered unsatisfactory by some stakeholders. We note that the importance of accountability within the regulatory framework was highlighted by the Department for Business, Innovation and Skills[1]. Establishing a

[1] Department for Business, Innovation and Skills (2011) *Principles of Economic Regulation* (link)

transparent set of principles in advance of the price review will enhance the predictability and consistency of economic regulation and so ensure that the water sector continues to be attractive to investors, while delivering an excellent value service to customers. Additionally, the legitimacy of subjective decisions will be boosted if there is a transparent link between the outcome and its origin.

The following principles have been shaped by our reflection on cost assessment at PR19, the subsequent redeterminations by the CMA and what potential challenges lie ahead at PR24. We were generally supportive of Ofwat's approach to cost assessment at PR19, and consider it a significant improvement to PR14. We expect continuation of the debate between regulator and industry on where the best path for PR24 lies. We hope that this paper will be able to frame future discussions and help all parties move toward a mutually supported outcome at PR24.

This paper has been split into six sections, each aligned with a particular principle:

- **Principle 1** sets out why it is important to consider how the design of cost assessment is best able to reflect the services provided.
- **Principle 2** emphasises the importance of embedding engineering and economic rationale within cost assessment.
- **Principle 3** considers how the regulator can best align the package of incentives to those found in a competitive market.
- **Principle 4** examines how the characteristics of the industry steers the cost assessment methodology in certain directions.
- **Principle 5** stresses the need for a coherent approach to cost assessment.
- **Principle 6** sets out the basis for an objective framework to assess efficiency.

Principle 1: Define the services provided

Headlines

- The cost assessment framework should identify the services that utility network providers deliver, particularly where multiple services are provided through a single value chain.
- The same service provided under different operating conditions can lead to different levels of expenditure within each element of the value chain.
- Defining the services provided and interdependencies between value chains should be the first goal of cost assessment.

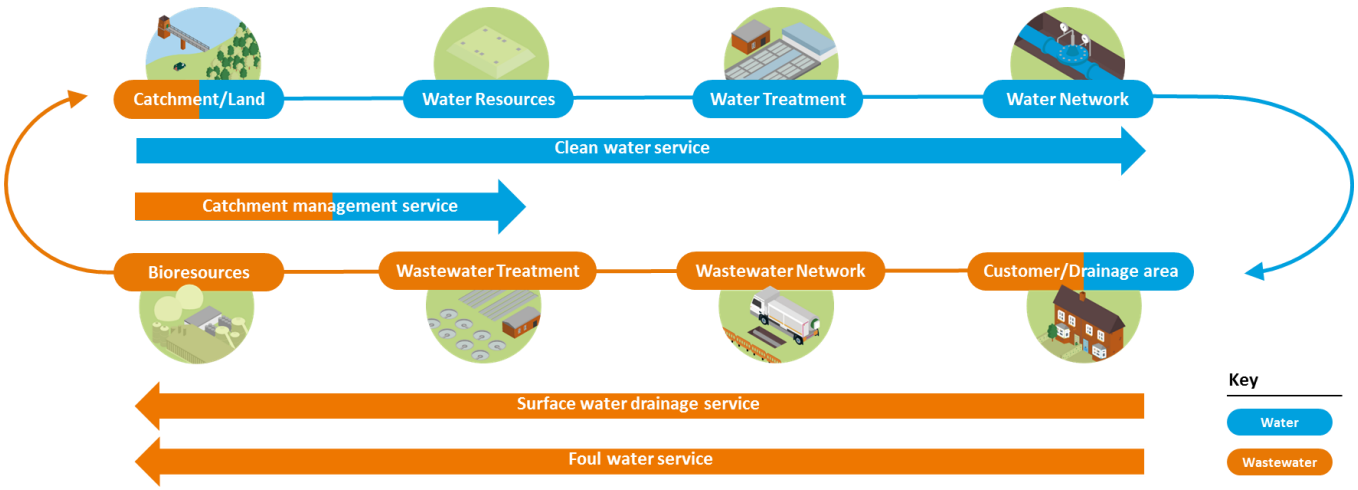
The primary aim of cost assessment is to identify the efficient costs for delivering a defined service. This means that the first step must be to understand the service(s) provided within each value chain. Only once this is fully understood can the regulator begin to develop the appropriate tools to estimate the cost of providing each service. Having a clear definition and understanding of the services provided enables us to explicitly target cost assessment towards an appropriate outcome.

Service definition is particularly important when considering the complexities of modern utilities. Each aspect of the value chain can deliver one or more distinct services. Importantly, companies will optimise their operations differently across the value chain in response to local operating requirements, when delivering the same service. This is known as a 'substitution effect'. This means that efficient spend within each service can vary across the industry in question, 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. Therefore, the **cost assessment framework should explicitly consider the services provided within the value chain**, and whether regional operating circumstances mean the efficient cost of delivering these services should vary between companies.

Identifying the services across each value chain

In this section, we illustrate the importance of identifying services within the water and wastewater system. A system is a set of elements that are interconnected and work together to serve a common function or purpose. It is important to understand how the different aspects interact with one another, but also how service offerings differ across the elements. Figure 2 illustrates the system within which water and sewerage providers operate as well as a simplified view of the core services provided across each element.

Figure 2 The services offered by companies as part of water and wastewater operations



Even from this simple illustration of the services provided within the system, we can see how different regional characteristics could lead companies to structure their operations differently (for example in response to water availability). In this way, different companies adopt different levels of activity across the system to deliver the same core services and improvements for customers. This is not to say that there are not any common factors that lead to comparable costs between companies; there are. However, the key challenge for cost assessment is to acknowledge that companies may have adopted different activity levels within each element for entirely legitimate reasons, and that this will lead to differences in efficient cost across the industry.

To meet this challenge, it is important to understand not just the costs associated with each service and within each element of the system, but also any interdependencies and trade-offs. Some of these interdependencies are clear in Figure 2, but it is equally important to understand where there are no connections. For example, if a company was able to reduce water consumption, this would require less water to be abstracted, treated and transported and it would result in less being collected, treated and returned to the environment. However, it would not have any impact on surface water drainage related costs across the wastewater value chain.

This highlights that foul and surface water collection are two distinct services, with different characteristics that vary for each element of the system. Foul sewerage relates to domestic customers and trade effluent. This service has more predictable costs, because the total load received from customers can be reasonably accurately predicted from year to year. Surface water drainage relates to the run-off into sewers caused by rainfall. By its nature, the costs associated with providing this service vary from year to year, simply because weather patterns vary from year to year and are uncorrelated to other exogenous drivers of cost, like population. Developing this understanding of the system and services offered enables *a priori* assumptions to be formed and the impacts of interventions to be appropriately assessed.

Principle 2: Prioritise engineering and economic rationale within cost assessment

Headlines

- Cost drivers should be justified on **engineering** or **economic** grounds.
- There should be a clear ***a priori*** relationship between the chosen set of cost drivers and observed cost.
- **Engineering** and **economic** rationale should be codified within the cost assessment framework using an objective model selection criteria.

Having identified the different services provided, the next stage of deriving an independent benchmark requires the regulator to identify factors that both drive cost within (and across) these services, and are capable of explaining efficient variations in cost for each company. Once causal narratives have been established[2], the regulator can consider the best way to reflect these within cost assessment.

This process should be built upon a deep understanding of the engineering and economic narratives[3] underlying the service provided. The intuition behind this derives from the statistical concepts of *causation* and *correlation*. Correlation simply describes the size of relationship between two different variables, whereas causation demonstrates that one variable is directly affected by the other. While these concepts are by no means mutually exclusive, we are less concerned about whether a cost driver *correlates* well with cost, and are more interested in understanding how cost drivers *cause* differences in cost.

Correlation alone (particularly in a small data set) may occur by chance, or due to partial correlation with a related variable. This means that an over-reliance on correlation may result in models that inconsistently (or inappropriately) capture cost differences between companies, leading to an inappropriate benchmark, or inadvertently explain efficiency differences between companies (rather than explaining cost differences). Therefore, cost assessment should establish cost drivers based on engineering and economic rationale before assessing what data is available. This will ensure that the final model suite is reflective of engineering and economic logic, which will increase the chance the benchmark is properly reflective of companies' circumstances. This approach contrasts with data mining, where variables are tested for statistical significance with no consideration given to the underlying engineering narrative. Such an approach is likely to lead to overfitting and an inappropriate benchmark.

This is why we emphasise strong engineering and economic rationale should be prioritised over other considerations, such as statistical performance. For example, whilst there will always be a compelling argument to include a cost driver that has both a strong engineering prior and a strong statistical relationship, the regulator should not completely discount a cost driver on the basis that it performs poorly statistically. It is still a genuine driver of cost; if the regulator is unwilling to use the driver within its econometric model, it simply means that it must look for another approach to reflect that factor within the wider cost assessment framework.

"Our emphasis is to develop models that make sense, with cost drivers that adhere to engineering, operational or economic rationale."
Ofwat (2018) *Cost assessment for PR19: a consultation on econometric cost modelling*

[2] For more detail on these steps and assessment of causal cost drivers in general, please see Arup and Vivid Economics (2017) *Understanding the exogenous drivers of wholesale wastewater costs.* Available here.

[3] We discuss the difference between engineering and economic cost drivers in Principle Three.

Using engineering priors to understand causation and cost drivers across each service and value chain will also highlight the potential for substitution effects that may need to be accounted for within the assessment e.g. by assessing several parts of the value chain together. Additionally, it can highlight where it may be appropriate to exclude costs from the model, for example, because they are not directly related to service provision or are outside of management control.

As noted in Principle One, companies will allocate their resources to optimise their operations differently to suit local operating requirements when delivering the same service. This means that there may be material differences in the level of efficient spend within each service across the industry in question. Pursuing causal narratives will ensure that the chosen cost drivers appropriately reflect efficient inter-company variation in costs within the benchmark.

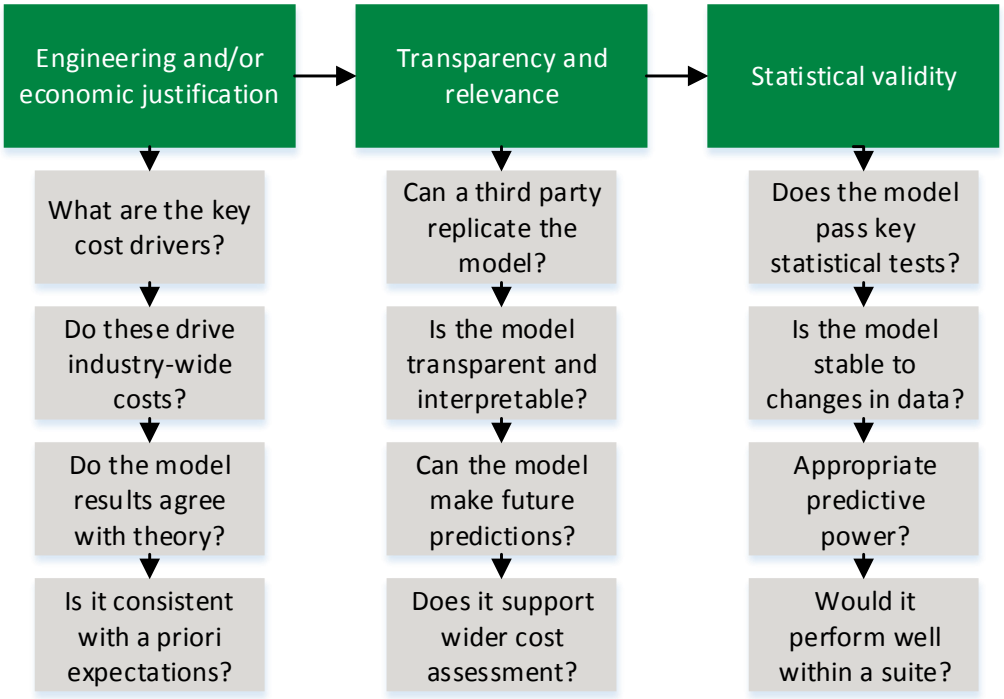
As we discuss further in Principle Four, **causal cost drivers will draw out robust relationships that can be used to assist in accurately forecasting expenditure in a future period**. Prioritising statistical performance and backward looking 'fit' will not provide the same assurance, as this only gives confidence in predictions for previous observations, not predictions of the future.

The importance of robust model selection criteria

Once engineering and economic rationale have been established, we can begin selecting variables and developing models. In this section, we provide an example of how we applied the principles set out in the previous section at PR19. Establishing model selection criteria *a priori* allows the model development process to focus on providing models that support the regulator's stated principles, as well establishing a transparent framework against which the regulator can objectively assess the performance of different candidate models.

During PR19, Ofwat held a consultation on econometric modelling during cost assessment. In their responses, water companies collectively proposed over 320 different models for Ofwat to consider. This exercise highlighted the importance of having a robust set of model selection criteria. The structure of the criteria should promote models with strong underlying engineering and economic narratives. For example, at PR19 we sequenced our model selection criteria deliberately to place the greatest emphasis on engineering and economic justification. This does not mean that we do not value statistical validity; this structure simply helps aid the assessment of models with strong engineering priors. This increases the likelihood that all appropriate engineering/economic drivers are accounted for within the process, either through modelling approaches (if it passes the subsequent tests) or through ex-post modelling adjustments.

Figure 3 UUW's model selection criteria at PR19



As Figure 3 illustrates, our starting point for model selection was the engineering and economic narrative underpinning a model. Once we established that a model is consistent with the relevant theory, we assessed the model's transparency and relevance as part of the second step. To understand the importance of this step, we need to remember that benchmarking is one of several components of cost assessment. Opaque models will mask any cross-over between the different components, possibly resulting in companies being compensated twice or not at all.

Additionally, we note that model replication is crucial if the regulator is seeking to provide a transparent outcome. Transparency helps to ensure that both companies and the regulator can interpret the results. The reality is that we are not going to develop one perfect model that works for every company for all future periods. Therefore, overly complicated or opaque models make comparisons to company plans more complicated as it becomes difficult to interpret what the baseline implicitly includes.

For the final step of our model selection criteria, we assessed the statistical performance of the model. Any model assessed at this point was grounded in engineering logic and complemented the wider cost assessment process. Applying statistical performance as a final assessment allowed us to choose the models most able to make robust forecasts of cost. We deliberately considered the statistical performance of a model last to discourage data fitting. Placing any emphasis on data fitting would dilute the focus on engineering narratives, which would objectively reduce the ability of the model suite to make predictions based on causal evidence (we develop this concept in Principle Four).

Testing the stability of the model by varying the measure's values using the reported confidence ranges will also enhance the robustness of the results. Variables that are reported with a large confidence/error range (for example, average pumping head) may lead to less stable, and therefore less credible results. Our PR19 model selection would drop such models/variables, and we would search for an alternative way of reflecting the cost driver within the cost assessment framework (either using an alternative variable or an out of model adjustment).

Principle 3: Protect the benchmark's independence

Headlines

- Cost drivers should be **exogenous** to company decision making, at least in the short term.
- Population cost drivers can be seen as **wholly exogenous**, while asset-based cost drivers are subject to a **degree of management influence**.
- Changing the **relative mix** of population and asset cost drivers will change the **exogeneity** within the model.
- Cost assessment **should not use endogenous** cost drivers, which are under immediate management control.

A key feature of a competitive market is that companies are price-takers. This means that no company is able to influence the price it charges, without reference to the prices in the market. This suggests that cost assessment will best mimic a competitive environment if regulated companies are less able to influence the outcome (for example, by manipulating cost drivers that are within company control). Additionally as noted above, the creation of an independent benchmark mitigates the problems caused by information asymmetry because it incentivises companies to reveal efficiencies and reduce costs to the efficient level.

For these reasons, we advocate that where possible, cost drivers should be exogenous to (company) management decision making. This approach maximises the incentive properties of the cost assessment process and most closely replicates the conditions of a competitive environment. This ensures that customers of regulated companies can benefit from excellent service quality at an appropriate price.

EXOGENEITY

- Exogeneity refers to where a company is unable to influence the outcome of cost assessment through its choices and actions.
- Exogeneity is desirable because it ensures the regulator's benchmark is independent of company plans.

During PR19, both Ofwat and the CMA recognised the need to focus more on using *exogenous* rather than *endogenous* drivers when seeking to develop a robust benchmark. We do however acknowledge that the water industry is asset intensive and so there may be circumstances when the use of a cost driver that is (at least partially) endogenous may be required. These are discussed in the next section.

Understanding limitations of exogeneity

An entirely exogenous cost driver is often difficult to find, particularly when dealing with an asset intensive industry, a small number of companies and relatively complex operations that vary regionally. However, some cost drivers can be exogenous in the short-term, but under management influence over the long-term. We refer to these cost drivers as short-term exogenous drivers, which aligns to the concept of an 'engineering' cost driver discussed in Principle Two. In contrast, truly exogenous cost drivers tend to be related to the population served or geographical characteristics of the region, for example the number of properties served by the company. We refer to these cost drivers as long-term exogenous drivers, which aligns to the concept of 'economic' cost drivers discussed in Principle Two.

We note that the difference between these drivers is not always clear-cut, particularly where management action is prompted by the characteristics of the region served. For example, a company may choose to build more pumping stations to deal with a hilly region. The challenge for cost assessment would be to assess whether this management decision was unavoidable and in response to exogenous impacts.

Short and long term exogeneity should be considered during the search for cost drivers. We would advocate that long-term exogenous drivers should preferred over short-term exogenous drivers, unless there is a compelling engineering, operational or economic justification for favouring a specific short-term driver.

We caution against the use of cost drivers that are not exogenous in the short-term or long-term. We refer to these cost drivers as endogenous drivers. Endogenous means subject to management influence. We draw upon ongoing debate in the water industry to provide an example of an endogenous cost driver below.

Exclude performance from cost benchmarking models

The most commonly suggested endogenous cost drivers are measures of performance, for example levels of leakage. Whilst these might legitimately result in differences in cost, we do not consider that benchmarking models are the best mechanism through which to compensate higher service quality for the following reasons:

- The inclusion of a performance driver would allow regulated companies to influence the benchmark through management decisions. This does not align to the concept of a price-taker discussed above[4].
- There would be a risk of duplicating (outcome delivery) incentives relating to performance. This does not align to the incentives found in a competitive market, which violates the economic rationale principle.
- The relationship between performance and cost is not always clear. This failure to establish clear *a priori* expectations of the relationship violates the economic rationale principle.
- Future costs of managing performance may be significantly different due to innovations that may not be appropriately reflected by the more general efficiency adjustment.

Instead, we consider that alternate mechanisms are better suited to target the potential for additional marginal costs of performance and complement the benchmarking process. For example, if a company believes that its service performance causes it to operate with comparatively higher costs, then it should seek that additional allowance through the submission of a well-evidenced cost adjustment claim (see Principle 5).

This protects the integrity of the independent benchmark while providing a company with an avenue to add efficient spend to its baseline. This approach is complemented by the Outcome Delivery Incentive (ODI) regime, which rewards/penalises companies for good/poor performance. We consider this structure to most closely replicate the incentives faced by companies within a competitive market and that cost benchmarks should not include such assessments.

INCENTIVISING PERFORMANCE

The regulatory regime for Water across England and Wales provides two main ways for companies to seek remuneration for the cost of providing a higher quality of service:

1) The Outcome Delivery Incentive (ODI) regime. This provides financial and reputational incentives for companies to achieve a pre-agreed level of service.
2) The cost assessment process (cost adjustment claims). This allows companies to request additional cost allowance in cases where the benchmark models do not capture the higher costs faced by a company.

[4] There are additional technical reasons underlying this statement: econometric theory states that the inclusion of an endogenous cost driver will result in a biased model. This would undermine the model's external validity (see Principle 4) and cast doubt on future spend forecasted by the model.

Principle 4: Ensure efficient expenditure outside of the modelled historical period.

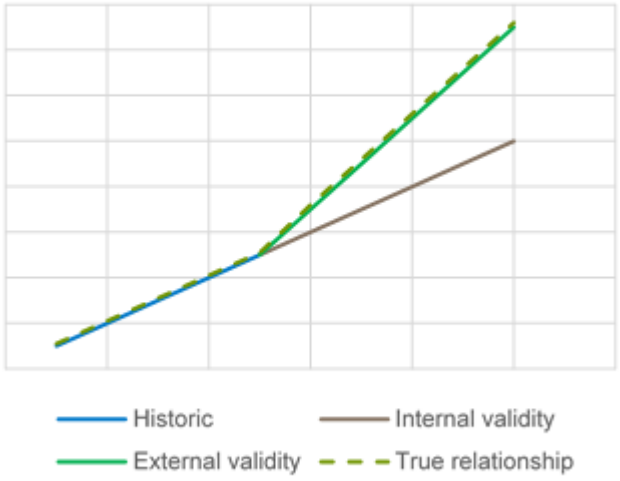
Headlines

- Emphasising **external validity** ensures that the model suite will be able to **forecast** expenditure in the next regulatory period.
- In contrast, placing emphasis on the **internal validity** of the model suite prioritises statistical significance.
- **The two are not mutually exclusive** and cost assessment should draw upon both approaches.

Cost assessment should recognise all services within a company's past and future operations and key operating and environmental differences. This means that it cannot be a solely backwards looking assessment based on what has been delivered previously; it also needs to account for what might be different in the future.

This has implications for the methodology underlying cost assessment. It is not sufficient to simply extrapolate historic trends forward with no consideration given to the underlying causal narratives; doing so could result in a misguided benchmark. Instead, **it is as important to consider the external validity of the model**.

WHAT WE MEAN BY INTERNAL AND EXTERNAL VALIDITY



- **Internal validity** is where we can make a prediction that best fits historical trends; simply put, it is how well we can predict what has previously happened (in the sample period).
- **External validity** preserves the key relationship between cost and cost driver, but crucially accounts for any changes in operating circumstances, allowing predictions into future periods.
- Both approaches are **equally valid** and should be utilised within cost assessment, but the decision to focus on one should be driven by the defining features of the issue at hand.

Internal validity is often the more familiar concept to users and is typically the simpler of the two to evidence, through the use of statistical and robustness checks for example. However, all these techniques prove is that the model is capable of predicting what has previously happened, to a reasonable degree of accuracy. As stated in Principle Two, tests for statistical significance should only be used to select between models that have strong engineering priors, and should not be employed as an initial parsing tool. This means that internal validity is referred to as a final sense check, to identify the best possible model suite.

Within regulatory cost assessment, we value internal validity but advocate placing more weight upon external validity because of a defining feature of benchmarking within regulated industries: small datasets. This issue intensifies the trade-offs already inherent within cost assessment and complicates the choice of an optimal set of cost drivers and the need for out of model adjustments. We have already stated that we consider the search for, and inclusion of, causal cost drivers to be crucial in defining an appropriate benchmark. This is because a small dataset tightens the constraint within which a benchmarking model is able to make robust predictions. Simply put,

any preference for a methodology focused on internal validity is foreclosed because the dataset we depend upon for benchmarking analysis is not able to support it.

"Benchmarking models should be strictly tailored to engineering narratives and avoid techniques that produce unstable results with small data sets."
Arup and Vivid Economics (2017) *Understanding the exogenous drivers of wholesale wastewater costs in England and Wales*

Additionally, an approach focused on internal validity might lead the regulator to conclude that an apparent industry-level relationship can be applied to individual companies without further investigation. However, the variation in how companies choose to structure their operations in response to local factors means that this approach might result in a sub-optimal outcome. This is particularly the case given the likelihood that benchmark models are not reflecting all relevant cost drivers due to limited datasets, meaning that it would not be obvious to the regulator that further work is required to refine the relationship in question.

If we use sound engineering priors to develop models then we would expect the relationship to hold in the future too, offering a degree of external validity. However, simply rolling forward the models to the future implicitly assumes that the past is a good predictor of the future; if things change, then this assumption would not hold, and the external validity of the models would be objectively weakened.

We can increase the level of external validity by deriving accurate forecasts of the selected variables, by accounting for changes in population growth for example. However, other factors that influence a company's costs, which are not accounted for within the modelled variables (climate change for example), are often more difficult to address. If sufficient levels of external validity cannot be offered through accurate forecasting, then we should realistically be more cautious about the ability of the benchmark to accurately predict future requirements. In such instances, additional out of model adjustments might be required.

Measurement error

Measurement error is the difference between the measured value of something and its 'true' value. It can occur because of reporting inconsistencies between companies. A good example is the average pumping head measure reported by companies in their annual reports. Theoretically, it should be the best measure of regional topology but has not been favoured due to clear differences in company reporting. Measurement error can affect both the predictive power of models and the interpretation of results, and particularly affects disaggregated models, due to the presence of different cost allocation practices and substitution effects.

We note that it would be almost impossible to eliminate measurement error entirely, so its effects must be recognised within the cost assessment framework, notably the reliance placed on the modelled result versus the need to make out of model adjustments. We discuss this further in Principle 6 by reference to setting an appropriate efficiency challenge.

"Measurement error produces ranges in assessed company costs worth many hundreds of millions of pounds, which will be compounded by further error in the efficiency challenge."
Arup and Vivid Economics (2018) *The use of econometric models for cost assessment at PR19*

Principle 5: Ensure there is a coherent approach to cost benchmarking and the wider framework

Headlines

- The cost assessment methodology should **explicitly acknowledge** the degree of model parsimony it is targeting and the consequent approach taken to cost adjustments.
- A **diverse** model suite will reduce the need for a substantial number of cost adjustment claims, whereas a **parsimonious** approach will increase it.
- Theoretically, the chosen balance is irrelevant to the accuracy of the final outcome.
- As a general rule, botex and enhancement should be assessed separately within a coherent framework.

A key challenge faced by the regulator is ensuring that its cost assessment methodology is joined-up; it must be mindful that decisions in one area can have consequential impacts in another. For example, this is particularly true when considering how the cost adjustment and enhancement framework best complements the benchmarking approach.

We consider that the approach to cost adjustments should be driven by the approach taken during the benchmarking process. If benchmarking draws upon a wide variety of cost drivers and information, such that the expectation is that different company circumstances are reasonably represented within the benchmark, then the need for cost adjustments should be reduced (presuming the benchmark has external validity).

If, however, a more parsimonious approach to developing the benchmark is taken, then the consequences are that companies can be expected to submit a large number of cost adjustment representations to ensure that their individual circumstances are reflected in the benchmark. The remainder of this section develops this idea, and concludes by considering the principles that might guide methodological decisions to pursue simplicity or diversity. It also discusses additional considerations that can help ensure the wider framework is coherent with the benchmarking approach.

PARSIMONY

- A parsimonious model is a model that achieves a desired level of goodness of fit using as few explanatory factors as possible.
- Parsimonious models are usually easier to interpret and have more predictive ability, as they are less likely to overfit the original dataset.
- The choice of approach should be driven by the characteristics of the dataset used within cost modelling.

The importance of an objective adjustment framework

Accepting that modelled baselines can be imperfect at a company level necessitates a framework that allows companies to represent on potential reasons for their efficient level of cost to differ from the benchmark. Within the water industry, these are known as cost adjustment claims.

"Cost adjustment claims are mechanisms for a company to present evidence of unique operating circumstances, legal requirements or atypical expenditure which drive higher efficient costs for the company relative to its peers."
Ofwat (2019) *Initial Assessment of Plans, Technical Appendix 2: Securing Cost Efficiency*

Cost adjustment claims may be necessary where an approach (for example a statistical model suite that utilises a small dataset) is unable to adequately predict all sources of differences in efficient cost between companies. Conversely, if it were possible to develop a statistical model that is able to perfectly predict every company's efficient spend, then no cost adjustment mechanism should be needed. **This leads to the understanding that there is a direct relationship between the approach to deriving expenditure allowances and the need for cost adjustments**.

It is important for the regulator to consider the interdependencies between its benchmarking approach and the extent to which this will need to be supplemented by cost adjustments. Parsimony is often discussed in a negative context but it is an entirely legitimate methodological choice, particularly when dealing with a small dataset and the need to be transparent. However, incumbent with this approach comes a much higher likelihood that an individual company's benchmark might not fully reflect the drivers of efficient cost. For this reason, **the evidential bar applied to cost adjustments must be inversely related to the level of parsimony pursued by the regulator** i.e. the more parsimonious the initial benchmark, the more accommodating the regulator needs to be of cost adjustment claims.

We can quantify the degree of parsimony using various quantitative approaches[5] but it should primarily be something that is assessed through qualitative assessments of the services and engineering rationale that have been identified at the start of the process and how well these are accounted for within the overall suite of models. While useful, quantitative approaches to assessing parsimony aren't able to assess the parsimony of the model *suite*, and instead focus on an *individual* model. An individual model might be more parsimonious and therefore unable to account for all of the differences between companies. However, if it is utilised as part of a diverse model suite then the ability to make an accurate prediction for all companies is improved, in effect reducing the parsimony of the result and potentially reducing the need for further adjustments.

Figure 4 the relationship between the degree of parsimony in the benchmark modelling and the need for a cost adjustment process

Degree of parsimony

1	2	3	4	5	6	7	8	9	10
10	9	8	7	6	5	4	3	2	1

Cost adjustment evidential bar

This approach should not be interpreted as an avenue through which companies can seek inefficient additional funding. Rather, it recognises the fact that while statistical noise may 'cancel out' at an industry level, this is not necessarily the case at a company level. Therefore, we stress the need to supplement an industry benchmark with objective company-specific adjustments, which will allow the regulator to strike an appropriate balance within the cost assessment framework and provide a legitimate outcome for all stakeholders. We note that an objective framework to develop an independent benchmark would be applied without reference to preconceptions of what a 'reasonable' level of spend is. If the outcome is to be considered legitimate, any cost adjustment claim that meets the evidential threshold should be included within the view of efficient cost.

Ensuring that the approach to benchmarking is coherent with the wider regulatory framework.

Regulators typically prefer to assess the majority of costs using benchmarking techniques. The methodology followed during this process has a material impact on the allocation of benchmarked costs across the industry, but it also has implications for the wider cost assessment framework. Benchmarking models are not able to provide a granular allowance for the activities included within them; they simply provide a 'top-down' estimate. This creates

[5] Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) or Minimum Description Length (MDL) being the most commonly used approaches

difficulties when stakeholders are trying to understand which activities have been funded, and to what extent, particularly relating to enhancement activities. As a rule, we consider it more sensible to assess enhancement costs separately from botex. This is partly because the 'lumpiness' of enhancement expenditure is not suited to the relatively static cost drivers generally used in benchmarking models. Assessing enhancement and base expenditure separately also makes the prediction much more transparent, and reduces the need for stakeholders to make subjective calculations of 'implicit allowances'. Because of the difficulties created when assessing botex and enhancement together, any change to the methodology in this regard should be consulted on with stakeholders, and the consequences on other aspects of the price control carefully thought through before any change is made.

While coherency of the cost assessment framework is an important end in of itself, it is equally important that decisions taken within cost assessment also complement the wider regulatory framework, either directly or indirectly. Outputs from cost assessment are directly used in a number of areas, such as the derivation of grants and contributions, and as an input into financial modelling to calculate allowed revenues.

These interdependencies should be identified at the outset and cost assessment should be constructed to ensure that these outputs could be generated accurately and transparently. Indirectly, cost assessment should be set in a way that complements the outcomes regime and delivers wider regulatory ambitions (promoting long-term resilience for example) through providing appropriate incentives. Within outcomes in particular, there is a need to ensure that companies are not remunerated twice for the same activity (once through cost assessment and once through an ODI) but conversely, that expectations of service and stretch opportunities are realistic and appropriately remunerate the company. In isolation, stretching performance and cost might look achievable but it is important that the two be assessed together in the round, to avoid setting unrealistic frontier challenges.

Additionally, it is unreasonable to assume that everything can be known when making a forward-looking prediction. Therefore, there may also be a need to account for uncertainty through the creation of adjustment mechanisms. Again, this will necessitate that we make cost assessment predictions with the appropriate level of granularity to facilitate the creation of an appropriate and transparent mechanism to ensure that both customers and companies are appropriately protected. If elements are more uncertain, then it might be more appropriate to assess them separately rather than including them within more aggregated approaches to benchmarking.

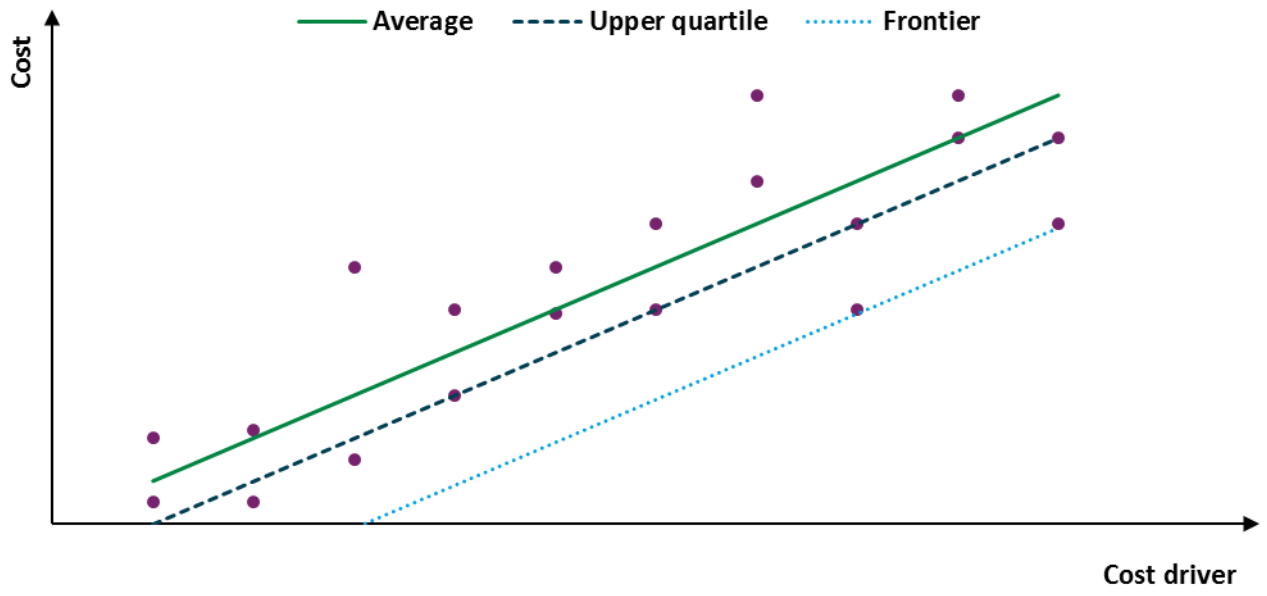
Principle 6: Challenge efficiency with a transparent, objective and stable framework

Headlines

- Efficiency should be challenged within a **transparent, objective and stable framework**.
- The efficiency target should set expectations that are **stretching but realistic**.
- The modeller should give due weight to the **impact of measurement error** on results.
- Efficiency challenges should **not** be set at a **disaggregated** level.
- **All sources** of efficiency challenge within a regulated industry should be **explicitly** considered.

Challenging efficiency

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 an objective framework if the aim is to produce an outcome that all parties view as legitimate.



Benchmarking models show the costs predicted by the chosen set of drivers, which are then compared to actual observed costs to infer efficiency. In reality, any unexplained variation between companies' costs is a mixture of companies' relative efficiency on one hand and company-specific noise and model biases on the other. Crucially, the models calculate 'average' efficiency, which can be thought to be relatively reliable because that is what econometric models are designed to assess. As the efficiency challenge is increased beyond the average level, the benchmark assessment becomes more reliant on the way that models assess the distribution of company performance around that "average cost" assessment. Absent a large number of data points, this distribution becomes less reliable as we move further from the average, which objectively reduces the reliability of tougher efficiency targets. Without any compensation within the efficiency assessment framework, the efficiency challenge might be unduly influenced by statistical noise (or anomalous performance of individual companies). This problem would be compounded by an efficiency challenge set at the value chain level, which would ignore the presence of substitution effects and so would risk setting an unachievable benchmark.

Methodological choices made earlier in cost assessment can influence the variance of the unexplained portions of cost. For example, a more diverse model suite should be associated with a better overall fit, and less unexplained cost variation. However, the corollary of this is that there would be a reduction in outlier values, and therefore an efficiency challenge that might be considered less stretching by a casual observer. This is not the case and is simply evidence that the benchmarking models are predicting costs effectively and so performing well, by removing noise from the residual and not from the underlying differences in efficiency.

Because of this, an objective framework for assessing efficiency should not seek to make ex post alterations in response to a perceived level of stretch that is different to prior expectations. Rather, the regulator should be fully cognisant of how legitimate methodological choices can lead to a reduced spread of residuals and so (entirely reasonably) a different catch-up challenge. Therefore, it should state the level of efficiency that it will set by reference to the percentile at the outset and maintain this assumption throughout the process. Stating that models might change because of e.g. additional data, is not a valid reason to change the percentile challenge. We would also caution against deviating from a framework that has been established ex ante, without giving stakeholders the opportunity to properly represent on the proposed change. Such action would undermine the legitimacy of the outcome. This is equally applicable to all areas of cost assessment.

An 'in the round' assessment

Ultimately, the efficiency challenge is largely a process of regulatory judgement rather than the product of a precise calculation. However, the result obviously has to be credible and transparent, otherwise companies will legitimately seek to challenge this. The assessment should account for all sources of efficiency challenge within the regulatory framework to ensure a coherent and legitimate outcome. It is important for the assessment to identify and quantify the total level of direct and implied challenge. Doing this deliberately and explicitly will provide all stakeholders with comfort that the outcome is stretching, but achievable and so demonstrate the legitimacy of the outcome.

Table 1 illustrates the different routes through which the regulator is able to challenge efficiency within regulatory cost assessment.

Table 1 The different types of efficiency challenge that can be placed on companies

	Source of efficiency	Comment
1	Catch-up challenge	The regulator applies a catch-up challenge by setting an industry-wide benchmark, usually by reference to a percentile or company for historical (modelled) expenditure.
2	Frontier shift / productivity challenge	This challenge ensures thresholds remain at the chosen frontier as companies may become more productive over time, due to learning, innovation or technological progress.
3	Relative price effect (RPE) allowance	The regulator can provide additional allowance or challenge for specific items that significantly contribute to a company's cost but that are outside of the company's control.
4	Restriction of allowance for service improvements	The regulator may choose to set expectations for service level improvements without providing additional cost allowances. This is in effect an additional productivity challenge.
5	Efficiency implicit within inflation	Allowed revenues and costs within regulated industries tend to be linked to a measure of economy-wide inflation. However, inflation is itself a function of the balance between increasing input prices, and economy-wide productivity improvements.

We would also note that a legitimate efficiency challenge should only apply to costs that are reasonably under management control. Engineering and economic rationale can be used to determine the rationale behind a decision to apply (or not) stretch against any element of cost. We consider that this should be done explicitly to support transparency within the cost assessment framework.