Reckon LLP
Data analysis | Economic regulation | Competition law

The use of performance ratios when benchmarking water companies against common PCs

Paper prepared for Anglian Water - 28 September 2023

Overview and key points

At PR19, Ofwat placed emphasis on its assessment of upper quartile levels of performance when it used cross-company benchmarking to set the performance commitment levels (PCLs) for some performance commitments (PCs). There is wider regulatory precedent from Ofwat and the CMA for using the upper quartile as a benchmark for the costs or performance levels that can be achieved by a notional efficient company.

There is a question of whether the use of upper quartile benchmarks across costs and performance is too demanding. This is outside the scope of this paper, but it is something that we expect to receive further attention during the remainder of the PR24 process, in light of outturn data on costs and performance in recent years.

This paper focuses instead on some more technical matters relating to the application of the upper quartile (and similar benchmarks) in the context of benchmarking companies' performance on common PCs and setting PCLs (leaving aside questions about how best to compare performance across companies). We present our three main suggestions in the box opposite.

Our attention to the upper quartile is not intended to endorse the setting of PCLs at upper quartile performance levels. This paper is applicable to a wider range of options (e.g. PCLs based on the median performance level, or some point between the average and upper quartile performance).

1. We suggest that Ofwat and companies draw on the established practice from water company cost benchmarking of calculating "efficiency ratios", to calculate corresponding "performance ratios" for companies' observed performance. An upper quartile or other performance challenge can then be derived from the set of performance ratios across companies.

2. We suggest that these performance ratios are calculated by looking at each company's performance over several years (e.g. five years), to take account of year-to-year fluctuations in the observed performance of even well-run companies.

3. We suggest drawing on evidence of companies' performance (and performance ratios) that is aggregated across multiple PCs, to help guard against the risk that the overall set of PCLs would not be achievable simultaneously by an efficient and well-run company.

Contents

In addition to these sections, the endnotes provide some more technical details which expand upon some of the earlier material.

Introduction and context

Ofwat's price control framework involves financial incentives for water companies across a set of performance commitments (PCs). These PCs concern aspects of performance such as water supply interruptions, leakage, discharge permit compliance, sewer flooding and pollution incidents. They relate to the outcomes for customers and the environment that companies influence.

Under Ofwat's approach, the "performance commitment level" (PCL) represents a baseline or target level of performance for the relevant PC. Typically, a water company faces financial penalties if its performance in a given year is worse than the PCL for that year and receives financial rewards if its performance is better than the PCL.

At the PR19 price review, the PCLs for some PCs were set at a common level across companies while others were set at company-specific levels. For the PR24 review, Ofwat plans a greater role for PCLs that are set at the same level across companies.

At PR19, where Ofwat applied common PCLs across companies, it set these using the upper quartile (UQ) of company forecasts of the performance for the forthcoming price control period. Ofwat also carried out its own analysis of historical performance, both across companies and over time, to inform the PCLs.

This paper focuses on some relatively technical and narrow matters relating to the way that upper quartile (or similar) performance benchmarks might be calculated and assessed when setting PCLs.

As part of this, we provide an initial discussion of what we have called "performance ratios", which we hope is useful to Ofwat and water companies. This is not intended to be definitive and we expect scope for further iteration and refinement of the ideas presented here.

We leave outside the scope of this paper some broader issues, such as: (a) the most appropriate methods or models for benchmarking performance across companies; (b) the extent to which historical trends in performance should be extrapolated into the future; and (c) whether the upper quartile is itself a reasonable regulatory expectation or too demanding (this is something that we expect to receive further attention during the remainder of the PR24 process, in light of outturn data on costs and performance from recent years).

This paper was sponsored by Anglian Water. It grew out of a project in which Reckon provided Anglian Water with analysis and insight to support its PR24 business plan proposals for stretching but achievable PCLs. The issues covered in this paper may be of more general interest and we saw value in presenting them in a self-standing format. The ideas and views expressed in this paper are the sole responsibility of the Reckon LLP project team.

Section 2

Efficiency ratios and performance ratios

We draw on the concept of efficiency ratios from Ofwat's approach to the benchmarking of water companies' costs

The concept of an efficiency ratio (or efficiency score) has become an established part of Ofwat's practice of benchmarking water companies' costs, especially for what Ofwat refers to as base costs. We outline in the box opposite how the efficiency ratio is calculated.

This concept of the efficiency ratio is quite a flexible one: it can be applied to historical costs or forecasts costs; calculated over a single year's data or multiple years' data; and calculated at different levels of aggregation (e.g. for just a single model or for a suite of models involving different specifications and/or spanning different parts of the value chain).

While the terminology of the efficiency ratio is used, it is important to keep in mind that differences between companies in their efficiency ratios are not just reflective of differences in relative efficiency. They may be driven by other factors (e.g. limitations in the ability of the benchmarking models to account for exogenous cost drivers, or variations between companies in the time profile of their expenditure).

Ofwat uses the efficiency ratio as a means to calculate the catch-up efficiency adjustment which it applies to the modelled costs for each company when determining price control expenditure allowances. In setting catch-up assumptions, Ofwat and the CMA have sometimes used upper quartile values from the range of efficiency ratios across companies.

The spread of efficiency ratios can also be informative on the relative performance of alternative benchmarking models or methods. A wider spread indicates a model that does less well in explaining the variation in observed costs across companies.

Step 1: Define a model (or model suite) or method for benchmarking companies' costs, and the scope of costs to be benchmarked.

Step 2: Define the time period over which the efficiency ratio is to be calculated.

Step 3: Apply the model/method from step 1 to calculate a modelled or predicted level of costs for each company over the time period in step 2.

Step 4: For each company, calculate the efficiency ratio as the aggregate of the company's outturn/forecast costs over the selected time period divided by the aggregate of the modelled cost for that company over that time period.

Efficiency ratios can be expressed as a ratio or as a percentage. A lower ratio means better company performance in terms of its costs relative to the benchmark.

A corresponding concept of performance ratios can be used when benchmarking the performance of companies against common PCs

Drawing on the idea of the efficiency ratio, we see a corresponding concept of the "performance ratio" which can be applied in the context of benchmarking companies' performance against common PCs. See the box opposite (and also endnotes 1 and 2).

Differences across companies in the calculated performance ratio are likely to reflect a range of factors besides simply company/management performance (see page 11) and the performance ratios will reflect the extent to which these factors are accounted for in the method or model used to benchmark performance.

Ofwat's approach at PR19 did not involve much explicit consideration of the most appropriate methods or models to benchmark companies' performance against specific PCs; Ofwat did not base common PCLs on results from econometric models. For the type of approach used at PR19, we propose that the denominator in the performance ratio is the average of companies' performance. This definition of the performance ratio does not limit flexibility as to how the PCL is set (see endnote 1 for further explanation).

Step 1: Define a model or method for benchmarking companies' performance against the PC. For some PCs the method may simply be to compare performance across companies and expect the same level of performance from all of them. For others it might involve a company-specific benchmark derived from an econometric benchmarking model or a composite performance metric.

Step 2: Define the time period over which performance is to be assessed for the purpose of the performance ratio.

Step 3: Apply the model/method from step 1 to calculate a modelled or predicted level of performance for each company over the period in step 2. If the method is simply to compare performance across companies and expect the same level of performance from all of them, the modelled performance should be set as average performance.

Step 4: For each company, calculate the performance ratio as the average level of the company's outturn/forecast performance over the selected time period divided by the average of the modelled performance level for that company from step 3.

NB: If the PC metric is such that a higher value on the metric implies better performance, apply the transformation described in endnote 2 to the performance data before making the calculations set out in steps 3 and 4. In most cases this transformation is not needed, but it helps ensure that a lower performance ratio consistently implies better performance.

Performance ratios can be used to calculate PCLs based on benchmarking results

1. Calculate performance ratios across companies for the PC

See previous page for further information on this stage

2. Choose appropriate definition of benchmark company

This might for instance be the upper quartile company, median company or something in between

3. Obtain the performance ratio for the selected benchmark

For instance, we might find that the UQ company has a performance ratio of 0.82, which represents an 18% "performance challenge" (or notional company adjustment) relative to the average level of performance

4. Apply the performance challenge*

This might involve combining the performance challenge with a modelled performance level or some other projection of industry-average performance for the forthcoming price control period

* If the performance ratio has been calculated through a transformation of the PC metric data (see endnote 2) then before calculating the PCL after the performance challenge, reverse the transformation to convert back to the units of the PC metric.

We see several applications for performance ratios

It allows for upper quartile benchmarks (or median, etc) to be applied in conjunction with performance benchmarking approaches that use econometric techniques to model trends in performance over time and to draw on these trends to make projections into the future (e.g. modelling a time trend based on a historical time series panel dataset).

It allows for upper quartile benchmarks (or median, etc) to be applied in conjunction with performance benchmarking approaches (e.g. econometric models) that take account of exogenous differences between companies that are considered to affect performance - beyond differences captured in the definition of the PC metric.

The range of performance ratios across companies provides insight on the absolute and relative performance of the method or model used to benchmark performance across companies. A wider range indicates a method/model that does less well in explaining the variation in observed performance across companies.

It helps provide transparency on the scale of any performance challenge - i.e. the extent to which a notional efficient and well-run company is expected to out-perform the industry-average level of performance. And it allows for the performance challenge to be calculated on an aggregated basis across a set of PCs, as well as at the level of a single PC.

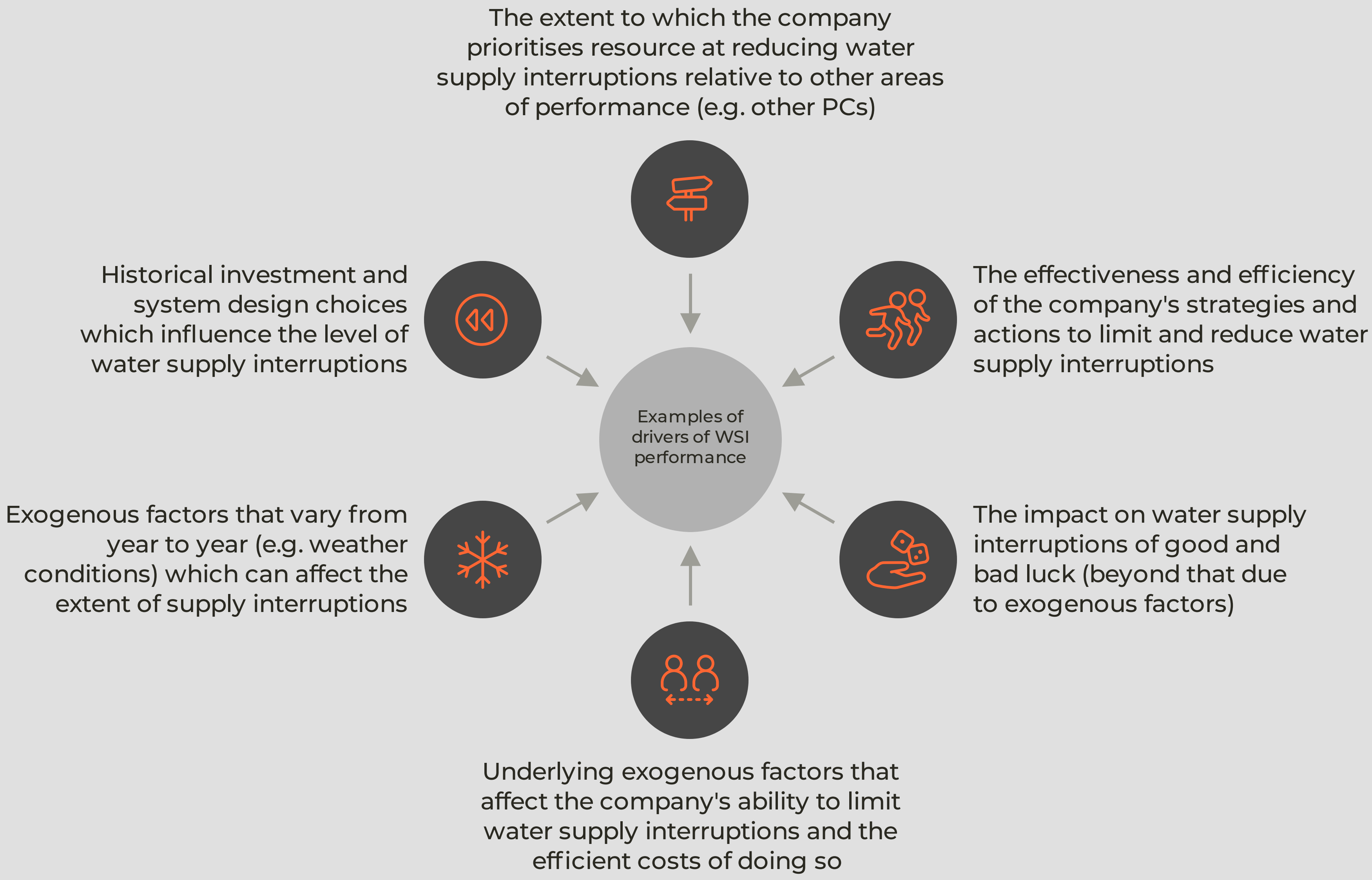
Sections 4, 5 and 6 of this paper consider in more detail the way that performance ratios can be calculated and used

Section 3

Factors influencing companies' performance

Before getting further into the application of performance ratios, it is useful to step back and consider factors that might give rise to observed differences in performance across companies and over time.

This diagram takes water supply interruptions as an illustrative example but the factors listed here have more general applicability.



Section 4

The time period over which performance ratios are calculated

As indicated on page 7, there is a choice about whether each company's performance ratio is to be calculated over a single year or based on an average of its performance over several years.

A performance ratio assessed using a single year's data would tend to be more heavily influenced by factors such as good or bad luck or weather conditions in the selected year, rather than reflective of the performance levels arising from companies' management and operational strategies. This, in turn, may make it unrepresentative of performance levels in future years - especially if a benchmark such as the upper quartile is used.

To take one example, performance in water supply interruptions can be affected by freeze-thaw events linked to the weather and by the extent of hot dry periods. While differences in companies' management and operational approach can affect the extent of weather-related impacts, even well-run and efficient companies may experience variations in performance linked to the weather.

When calculating benchmarks such as the upper quartile, for the purpose of setting PCLs for a forthcoming price control period, **we suggest that the performance ratio is calculated using performance data that covers a period of several years.**

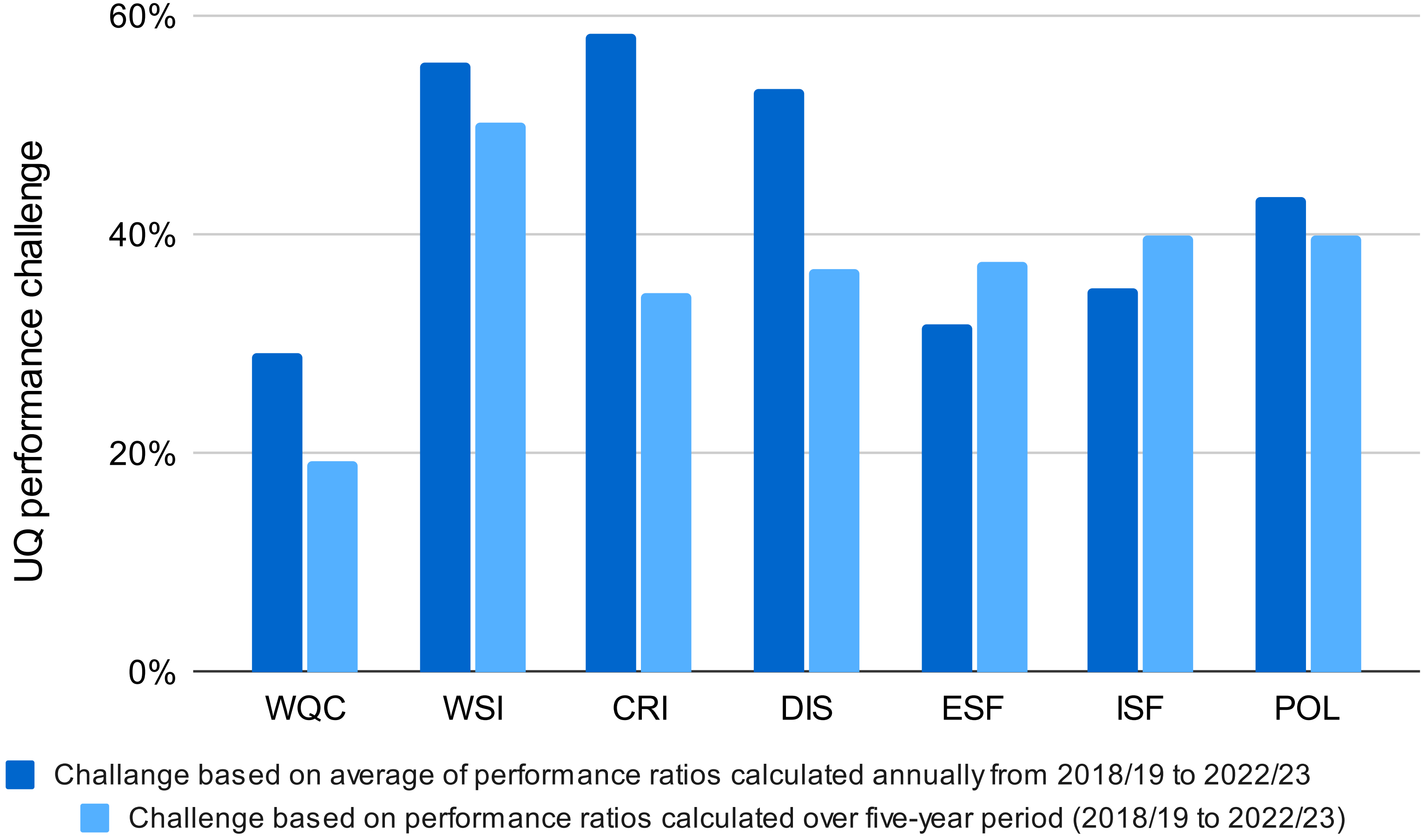
There seems to be particular merit in using a **five-year period**. This would help increase the likelihood that the calculated performance benchmark represents something which is achievable (in expectation) for a well-run and efficient company over the five-year price control period.

A five-year period is consistent with Ofwat's established practice from its approach to cost assessment. Ofwat has generally taken a five-year historical timeframe when calculating the efficiency ratios that are used to set the catch-up efficiency adjustments applied to modelled base costs.

Calculating the upper quartile performance ratio over a five-year period can either increase of decrease the scale of the implied UQ performance challenge, but it is more likely to reduce it where there is greater year-to-year fluctuations in each company's performance.

For each of the common PCs shown in this chart, the UQ performance challenge is calculated as one minus the performance ratio of the upper quartile company. The performance challenge represents the percentage reduction against average performance needed to achieve the calculated level of UQ performance.

Figure 1: Comparison of upper quartile performance challenge by calculation method - for selected performance commitments



Section 5

Combining performance ratios with time trend models

For PR24, Ofwat has signaled its intention to draw on analysis of trends over time in companies' performance when setting PCLs.

The methods for calculating and analysing trends over time in performance against common PCs has, so far, received relatively limited attention by Ofwat in its published documents. There are a number of different methods that might be used.

One key benefit of the performance ratio that we see is that it allows for benchmarks such as the upper quartile or median to be applied in conjunction with econometric modelling of trends in performance over time, as a means to inform projections of performance in the future.

In the chart on the next page we have taken the example of water supply interruptions and have calculated a historical time trend using an econometric model. For simplicity here, the model has a constant term and time trend only and is estimated using the same type of random effects approach Ofwat's uses for base cost benchmarking. Using this model, the time trend is calculated based on time series panel data covering the sample of water companies over the period 2011/12 to 2022/23.

We have extrapolated the time-trend into AMP8. The time trend from this model implies continuing performance improvements over time. Given the nature of the econometric model used, the modelled performance in AMP8 can be interpreted as a projection of something approximating industry-average performance in AMP8.

We have also calculated a projected upper quartile level of performance by multiplying the modelled industry-wide performance for AMP8 (based on extrapolation of the time trend from the econometric model) by the upper quartile performance ratio (calculated using performance data from the period 2018/19 to 2022/23).

A different way to model trends in UQ performance is to calculate an econometric time trend from a time series dataset of the *upper quartile level of performance* in each year (where the UQ performance is calculated separately each year and may come from a different company in each year). This trend does not use the full industry dataset.

Our view is that, of these two methods, the former is superior because it is based on an assessment of upper quartile performance that takes account of year-to-year fluctuations in performance even for efficient and well-run companies, and draws on industry-wide data to model the trend over time.

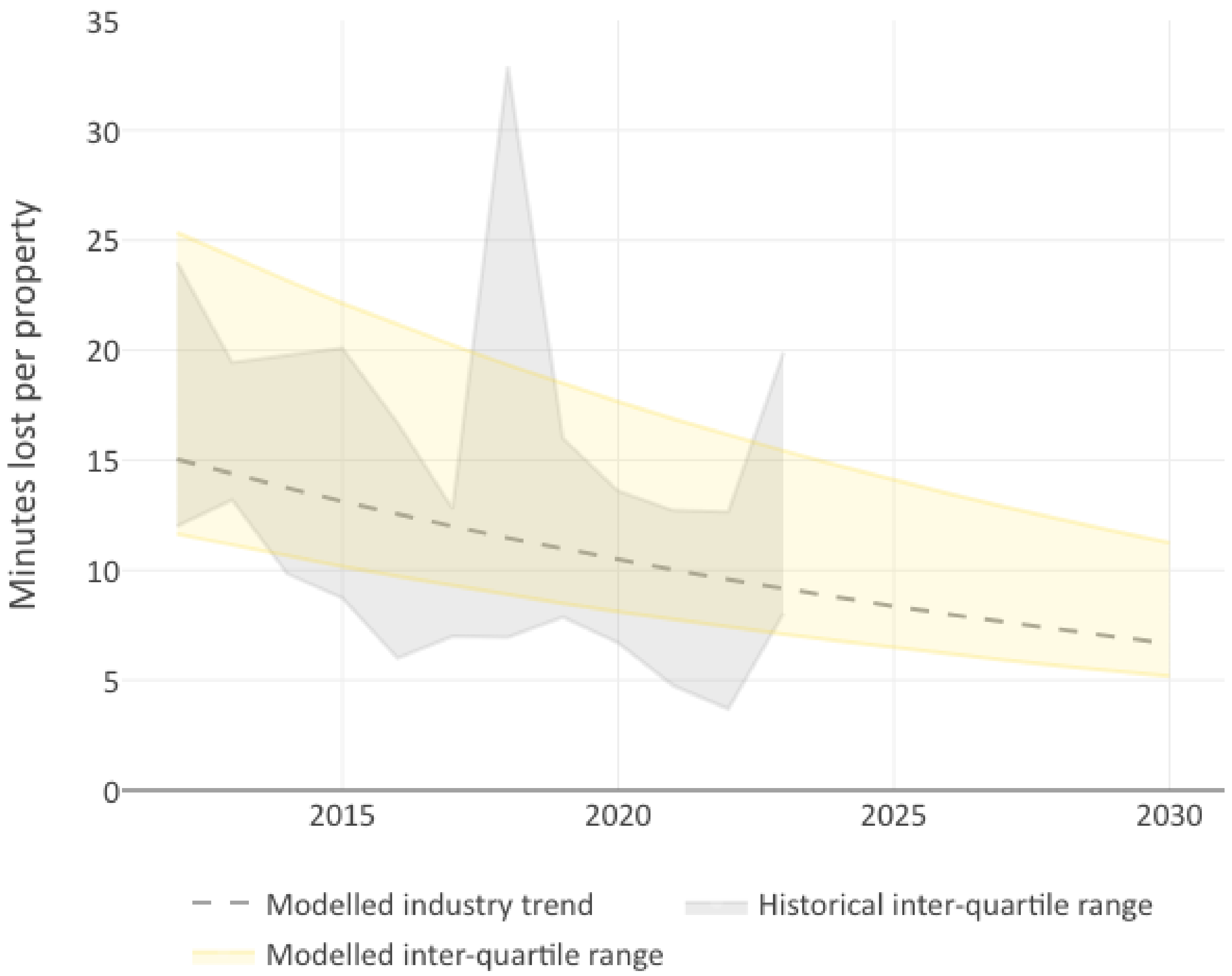
This illustrative chart shows modelling of a trend over time in water supply interruptions, with performance ratios used to calculate a modelled range between lower quartile and upper quartile performance projected into AMP8.

The yellow shaded area shows the modelled range between lower quartile and upper quartile performance, calculated by combining the modelled industry trend (estimated using historical data) with upper quartile and lower quartile performance ratios (based on companies' average performance over the period 2018/19 to 2022/23).

The grey shaded area shows the range between the upper quartile and lower quartile outturn performance in each historical year (these values are calculated separately each year).

The modelling approach and figures are illustrative only; there may be more appropriate ways to benchmark water supply interruptions across companies and over time (e.g. taking account of exogenous factors that can affect performance, or alternative model dynamics).

Figure 2: Modelling of water supply interruptions over time



Section 6

Performance ratios based on performance aggregated over multiple PCs

We see value in considering information on companies' historical performance at a more aggregated level than individual PCs

A further issue to keep in mind, when determining PCLs for a set of common PCs, is that it may be infeasible for a single well-run and efficient company to achieve an upper quartile (or similar) performance level simultaneously across all these PCs.

As highlighted on page 11, one of the factors that can influence a company's observed performance against a specific PC is the extent to which the company prioritises resource at improving performance under that PC relative to other areas of performance (e.g. other PCs). There may be significant differences across companies in terms of the allocation of attention, funding and other resources to individual areas of performance.

Insofar as the observed variation in performance across companies reflects such differences in management focus and resource prioritisation, then it may not be feasible for a single company to achieve upper quartile performance on each individual PC.

We do not consider it practical to try to strip out or adjust for any differences across companies in their prioritisation decisions when benchmarking performance. But it may be practical and illuminating to consider companies' relative performance - and to assess performance ratios - at an aggregated level across multiple PCs, rather than simply taking each PC in isolation.

There is some precedent for this from Ofwat's approach to cost assessment. Ofwat does not set catch-up efficiency challenges based on the upper quartile (or similar) efficiency ratio for each individual part of the value chain. Instead, for each price control, it aggregates results across the value chain (e.g. water resources plus and treated water distribution), before it calculates companies' efficiency ratios and before it determines the catch-up challenge.

The next page shows that, for an illustrative set of PCs, the group of upper quartile companies may vary by PC. The subsequent page discusses how performance ratios might be calculated at a more aggregated level than individual PCs.

There is variation across common PCs as to which companies are perceived as reaching upper quartile performance levels

The table opposite shows which wastewater companies might be identified as achieving upper quartile performance* (or better) within each of five individual PCs, based on the companies' average performance in the five years to 2022/23.

Given the set of 11 wastewater companies, there are three upper quartile companies for each PC. No company is upper quartile across all five PCs covered here. And seven of the 11 companies are upper quartile in at least one of these PCs.

We see a similar pattern on the water supply side: while there are some companies that are at the upper quartile or better for multiple PCs, there is variation across PCs in terms of the companies comprising the upper quartile group.

This type of evidence supports the view that an upper quartile performing company would not necessarily achieve upper quartile performance across each individual PC.

** See endnote 3.*

	Discharge compliance	External sewer flooding	Internal sewer flooding	Total pollution events	Sewer collapses
Anglian Water		UQ	UQ		UQ
Hafren Dyfrdwy					
Northumbrian Water				UQ	
United Utilties				UQ	
Southern Water					
Severn Trent Water	UQ	UQ		UQ	
South West Water			UQ		
Thames Water	UQ	UQ			UQ
Welsh Water					
Wessex Water	UQ		UQ		UQ
Yorkshire Water					

For the purposes of this paper, we are not advocating any specific method for calculating performance ratios that cover multiple PCs. We provide some examples of what might be done. There may not be a single best approach and this is an area that might benefit from further work in the future.

A simple way to look at companies' aggregate performance across a set of PCs is to first calculate the performance ratio individually for each PC and then take an average. However, there is not necessarily a good basis for giving equal weight to performance under each PC.

Another approach is to use information on ODI incentive rates. The logic for this is that, if we have performance metrics that measure different things (e.g. discharge compliance versus sewer flooding events) then we need some way to bring bring these into a common currency. ODI rates are used in practice to convert from performance against a PC metric to a performance penalty/reward in pounds.

For instance, a method that might be used is to first apply a common set of ODI rates to the performance of each company against each PC, to give notional £m performance values for each company and PC, and then to aggregate these £m numbers across PCs. An aggregated performance ratio can then be calculated from these aggregate £m figures. (See endnote 4 for further information.)

This could provide the basis for a sense check on the PCLs that Ofwat (or a water company) has determined based on a benchmarking approach that takes each PC individually.

For example, how does the upper quartile of the aggregated performance ratios calculated on the basis of companies' historical performance (under the ODI method above) compare with the aggregated performance ratio for a notional company that achieves the proposed PCL across each PC (using the same method to derive an aggregate performance ratio)?

If there is a significant gap between the two it could suggest that even an upper quartile company (operating across the set of PCs) would not simultaneously achieve the UQ performance levels for each individual PC.

Endnotes

Note 1

Where Ofwat simply compares performance across companies (without any specific model) - and implicitly assumes that each company would achieve the same level of performance if it was efficient - then under the approach we propose the "modelled performance" for each company would be the same and set equal to the *average* level of performance across companies..

The use of the average in these cases is to ensure that we have provided an unambiguous definition of the performance ratio. It would not be helpful in practice if the denominator in the performance ratio calculation was a subjective matter (e.g. a choice between say the average or median).

The definition of the performance ratio in this way does not limit flexibility as to how the PCL is to be set. For instance, the PCL could be set at a median or upper quartile performance level by setting PCL = modelled performance (i.e. average performance) * (performance ratio for the median or UQ company), which is analogous to the approach to applying a catch-up efficiency challenge in Ofwat's approach to base cost assessment.

Note 2

For some common PCs there is a positive upper bound to how good performance can be (i.e. 100% on the discharge compliance PC). For this type of PC, we propose that the value of each company's performance (and of average/modelled performance) used to calculate the performance ratio is taken as the gap between the company's performance and the upper bound (e.g. for a discharge compliance of 98.5% this would imply a performance value of 1.5%). This does not affect the ranking of companies but does affect the performance ratios calculated for each company and the spread of performance ratios across companies.

This transformation also has merit when using econometric models to estimate a trend in performance over time, in a context where performance over 100% is not feasible.

For other PCs, there may be no upper bound, but the PC metric is such that a higher value implies better performance. For these, we propose that the value of each company's performance (and in turn average/modelled performance) is calculated as one divided by the PC metric data for the purposes of calculating the performance ratio.

These transformations ensure that a lower performance ratio consistently means better performance across different PCs (but further refinement might be needed in some cases).

Note 3

The table is primarily for illustration and the cross-company benchmarking to identify the upper quartile companies is quite simplistic (i.e. based on the type of approach used at PR19 by Ofwat to set common PCLs). It does not necessarily reveal which companies are the best performers given their regional circumstances and other external factors.

Questions about the most appropriate methods or models to use for benchmarking companies' performance against common PCs are outside the scope of this paper.

Note 4

Where Ofwat's PCs are normalised by measures of company scale (e.g. sewer flooding events per 10,000 customers or pollution events per 10,000km of sewer) then the incentive rate (in £m) set for price control purposes will naturally be larger for companies operating at greater scale. But for the exercise here, it is important to use the same incentive rates for all companies, even if the incentive rates set by Ofwat vary by company. These might be based on the £m incentive rates for a single company or perhaps some cross-company average of the various incentive rates.

On this basis the £m performance value for each PC and each company represents a notional ODI penalty if the PCL was set at zero (i.e. zero pollution incidents, 100% compliance. etc). This is simply for the purposes of enabling aggregation in £m across PCs and does not imply anything about how the PCL would be set in practice. For the purposes of these calculations we also envisage the application of the transformations for PCLs such as discharge compliance where a higher value on the PC metric implies better performance (see endnote 2) before calculating the notional performance values.