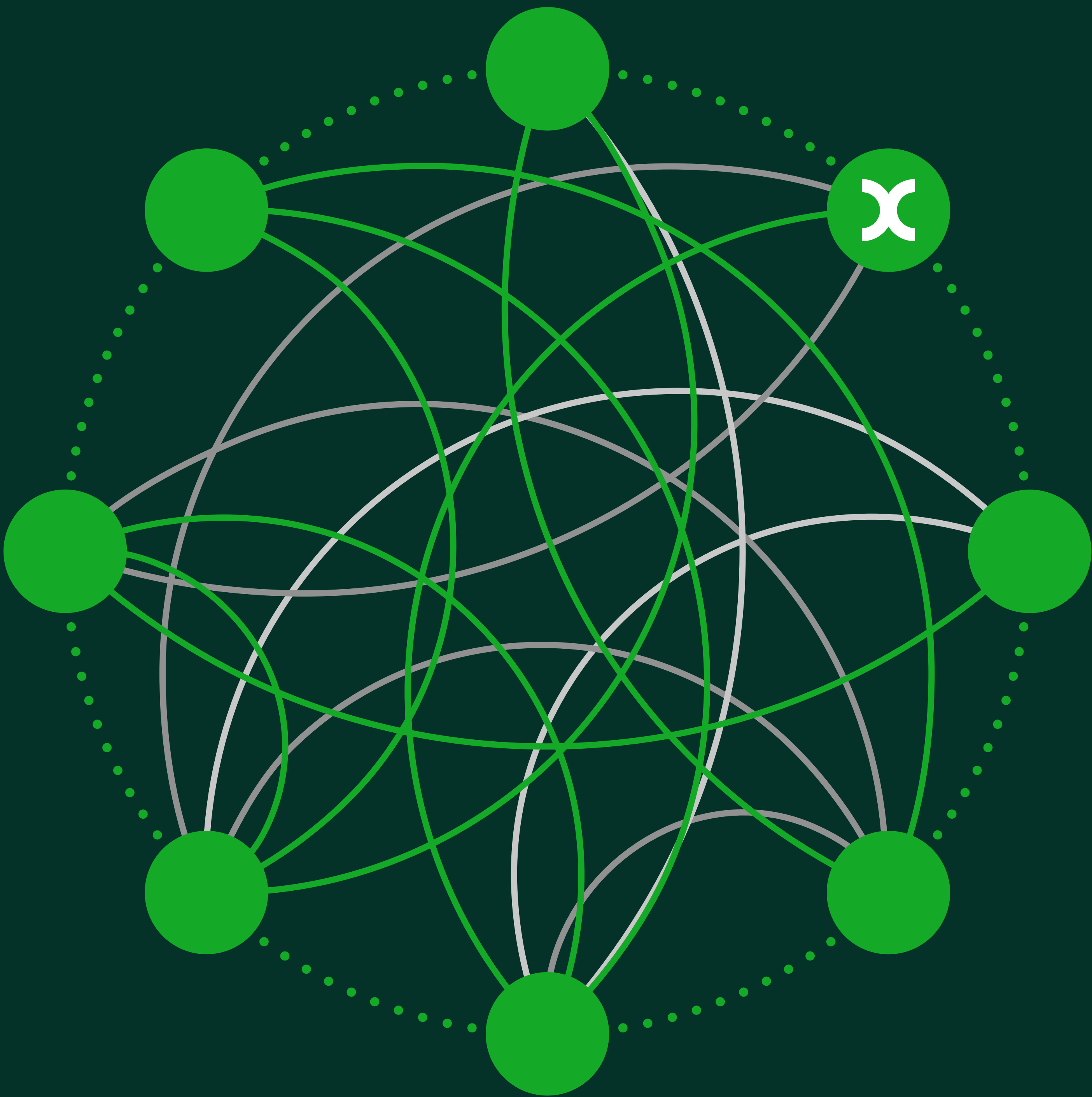
Econometric modelling of sewer flooding performance

—

Prepared for Dŵr Cymru

29 September 2023



Contents

Executive summary 1

1 Introduction 7

2 Methodology 13
2.1 The dataset 13
2.2 Defining the measures of service performance 14
2.3 Performance drivers 18
2.4 Model development process 24
2.5 Benchmark selection 25

3 Econometric models 27
3.1 Performance modelling of internal sewer flooding 27
3.2 Performance modelling of external sewer flooding 29
3.3 Performance modelling of total sewer flooding 31
3.4 Triangulation 33

4 Concluding remarks 36

Figures and Tables

Figure 1.1 Stylised example—models fund companies to achieve the same PCs 10
Figure 1.2 Stylised example—models fund companies to achieve the same PCs 11
Figure 2.1 Internal sewer flooding incidents (2019–23) 15
Figure 2.2 External sewer flooding incidents (2019–23) 16
Figure 2.3 Correlation between ISF and ESF (2019–23) 17
Figure 2.4 Total sewer flooding incidents (2019–23) 18
Figure 2.5 Distribution of rainfall drivers across the industry (2019–23) 20
Figure 2.6 Distribution of other performance drivers across the industry (2019–23) 22
Figure 3.1 Triangulation approach 34
Figure 4.1 TSF upper-quartile performance distribution: econometric versus Ofwat approach 36
Figure 4.2 ISF upper-quartile performance distribution: econometric versus Ofwat approach 37
Figure 4.3 ESF upper-quartile performance distribution: econometric vs Ofwat approach 38

Table 2.1 Correlations between performance and rainfall metrics 21

Oxera Consulting LLP is a limited liability partnership registered in England no. OC392464, registered office: Park Central, 40/41 Park End Street, Oxford OX1 1JD, UK; in Belgium, no. 0651 990 151, branch office: Avenue Louise 81, 1050 Brussels, Belgium; and in Italy, REA no. RM - 1530473, branch office: Via delle Quattro Fontane 15, 00184 Rome, Italy. Oxera Consulting (France) LLP, a French branch, registered office: 60 Avenue Charles de Gaulle, CS 60016, 92573 Neuilly-sur-Seine, France and registered in Nanterre, RCS no. 844 900 407 00025. Oxera Consulting (Netherlands) LLP, a Dutch branch, registered office: Strawinskylaan 3051, 1077 ZX Amsterdam, The Netherlands and registered in Amsterdam, KvK no. 72446218. Oxera Consulting GmbH is registered in Germany, no. HRB 148781 B (Local Court of Charlottenburg), registered office: Rahel-Hirsch-Straße 10, Berlin 10557, Germany.

Table 2.2	Correlations between performance and other drivers	23
Table 3.1	ISF modelling	28
Table 3.2	ISF models predicted performance levels for WSH	29
Table 3.3	ESF modelling	30
Table 3.4	ESF models predicted performance levels for WSH	31
Table 3.5	TSF modelling	32
Table 3.6	TSF models predicted performance levels for WSH	33
Table 3.7	Triangulated performance levels for WSH	35

Executive summary

Ofwat sets targets for both the efficient cost level and service quality metrics (e.g. performance commitments, PCs). At PR19, it used econometric modelling to set efficient base cost allowances for the majority of companies' base expenditure. In this way, companies' base costs were benchmarked against the most efficient companies in the industry, once a set of relevant exogenous factors were accounted for. Given the heterogeneity in the water industry across a range of characteristics (e.g. scale, population density, topography), regional factors need to be accounted for to ensure like-for-like comparisons between companies.

Meanwhile, Ofwat set several PCs using less sophisticated comparative benchmarking tools that did not normalise for relevant exogenous factors. For example, PCs on internal sewer flooding (ISF) incidents were normalised for scale and based simply on the upper-quartile benchmark performance. As relevant exogenous factors were not accounted for, there is a risk that these targets would not be achievable for companies that operate in complex environments that are more prone to sewer flooding (without additional funding), while the targets might be lenient for companies that operate under less challenging conditions. We understand that Ofwat has not provided an explicit justification for its approach, but has stated that it could see no 'clear reasons why companies should not achieve the same stretching levels of performance',[1] and that these metrics are 'both important to customers and supported by good quality comparative information'.

At PR19, Ofwat expected companies to meet these PCs through its base cost allowances, even though the PCs did not account for regional factors and the base cost models did not account for service quality performance. At the PR19 redetermination, the Competition and Markets Authority (CMA) provided additional allowances to some of the appellants to account for the costs associated with meeting the PCs on some service measures.[2]

[1] Ofwat (2019), 'PR19 final determinations: Delivering outcomes for customers policy appendix', December, p. 19.
[2] For example, see Competition and Markets Authority (2021), 'Anglian Water Services Limited, Bristol Water plc, Northumbrian Water Limited and Yorkshire Water Services Limited price determinations Final report', March, para. 4.495 and table 8.8.

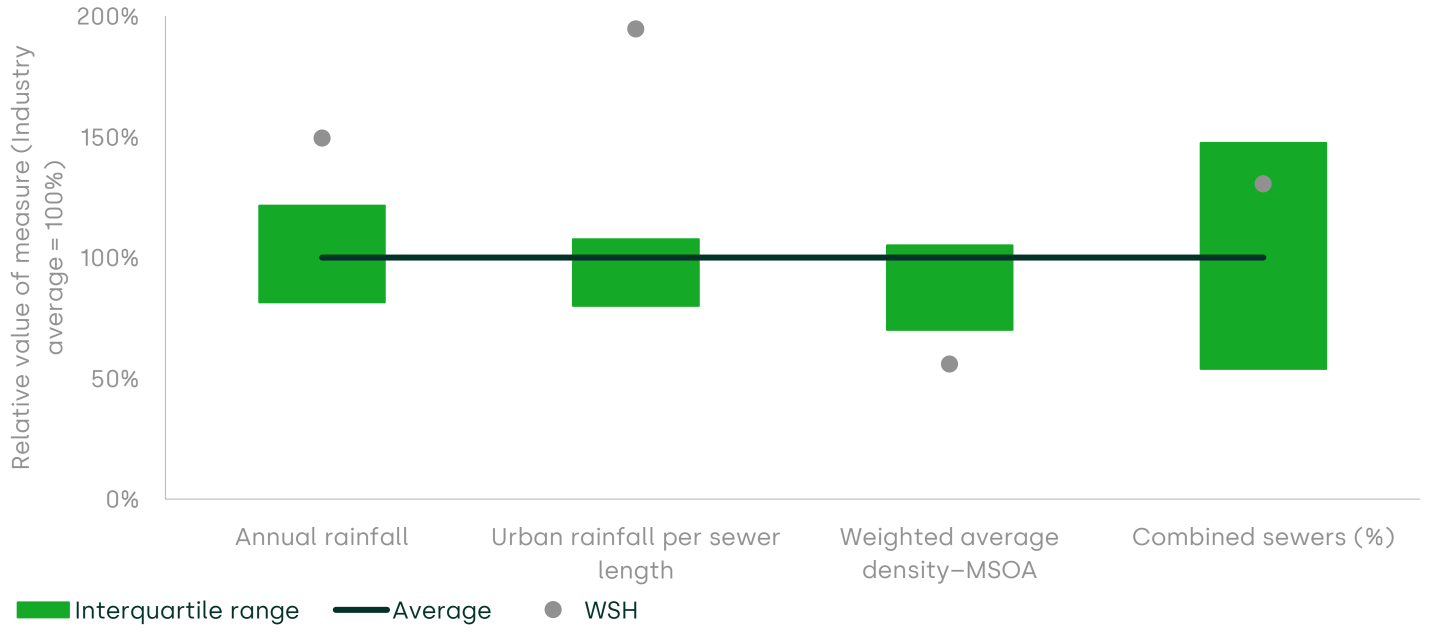
Dŵr Cymru (Welsh Water, WSH) has commissioned Oxera to identify key drivers of sewer flooding performance (both ISF and external sewer flooding, ESF), and to normalise for these drivers in an econometric framework consistent with Ofwat's approach to assessing efficient base costs. We also draw out implications from the econometric analysis for how these normalised performance targets ('econometric targets') compare with Ofwat's approach of benchmarking at the upper-quartile level without normalisation ('non-normalised targets').

ISF is generally defined as when water escapes from the sewer network and floods a main building, while ESF is defined as when the flooding occurs in the curtilage of the main building. While Ofwat provides some guidance on how companies should allocate flooding incidents between ISF and ESF,[3] there are incidents where a degree of judgement is required.[4] Therefore, as part of this report we explore total sewer flooding (TSF) models, alongside separate models for ISF and ESF.

Based on conversations with WSH, industry-wide discussions[5] and our own empirical investigation, we consider that sewer flooding incidents may be driven by several operational characteristics, including rainfall, population density and network characteristics. The chart below shows how WSH compares with the rest of the industry on some of these metrics.

[3] A description of ISF and ESF can be found in Ofwat (2018), 'Reporting guidance – Sewer flooding', https://www.ofwat.gov.uk/wp-content/uploads/2018/03/Reporting-guidance-sewer-flooding.pdf, accessed 10 August 2023.
[4] For example, in the case of a flooding incident on a farm, companies are required to provide a 'reasonable allowance' for what would constitute a garden (in which case, it is in the curtilage of the building and is considered ESF) and what is not (in which case, the incident is outside of the curtilage and is not reported).
[5] There is overlap between the performance drivers outlined in this section and the performance drivers considered in United Utilities' report for the future ideas lab on a similar topic. See United Utilities (2022), 'What lessons can we learn from cost assessment at PR19?', section 2.5.

WSH's position on performance drivers relative to the industry (2019–23)



Note: MSOA refers to Middle Super Output Area.
Source: Oxera analysis.

The chart shows that WSH operates in a relatively complex environment compared with the rest of the industry—it incurs significantly more rainfall (both urban and total), operates in a sparsely populated environment, and has more combined sewers than the industry average. Given its position on these three metrics, WSH may be less able to achieve non-normalised sewer flooding targets than the rest of the industry.

We note that Ofwat has accounted for some of these factors (e.g. population density) in its econometric cost models.[6] It has argued that adjusting PCs on the basis of regional factors that are already accounted for in the base cost models may result in a double-counting.[7] However, this argument assumes that there is no correlation between regional factors and service performance—an assumption that can be tested empirically, which Ofwat has not presented so far. Indeed, the argument is inconsistent with Ofwat's own approach to setting PCs, which, as noted earlier, normalises for one operational characteristic

[6] See Ofwat (2023), 'Econometric base cost models for PR24', April.
[7] See Ofwat (2022), 'Creating tomorrow, together: Our final methodology for PR24 Appendix 9 – Setting expenditure allowances', December, p. 61.

(scale) even though that operational characteristic is already incorporated in the cost models. Therefore, Ofwat's concerns regarding double-counting require further justification.

To normalise for these characteristics, we have developed models largely following Ofwat's cost modelling criteria (e.g. ensuring robustness from an operational, economic and statistical perspective). The statistical quality of these models is similar to that of comparable models presented by Ofwat at the PR24 cost modelling consultation,[8] and the models therefore provide a good starting point in the assessment of PCs.

The table below shows how WSH performs in these models (the 'econometric approach') under the upper-tercile (66th percentile) and upper-quartile (75th percentile) benchmarks, compared with how it performs under a non-econometric approach at the upper-quartile benchmark. The latter can be seen as roughly equivalent to Ofwat's approach to setting PCs.[9]

WSH's sewer flooding performance (2019–23): incidents per 1,000 properties

	ISF	ESF	TSF
WSH's outturn performance	0.16	2.72	2.88
Upper-quartile target: non-econometric approach	0.14	1.35	1.48
Upper-tercile target: econometric approach	0.17	2.45	2.72
Upper-quartile target: econometric approach	0.16	2.44	2.68

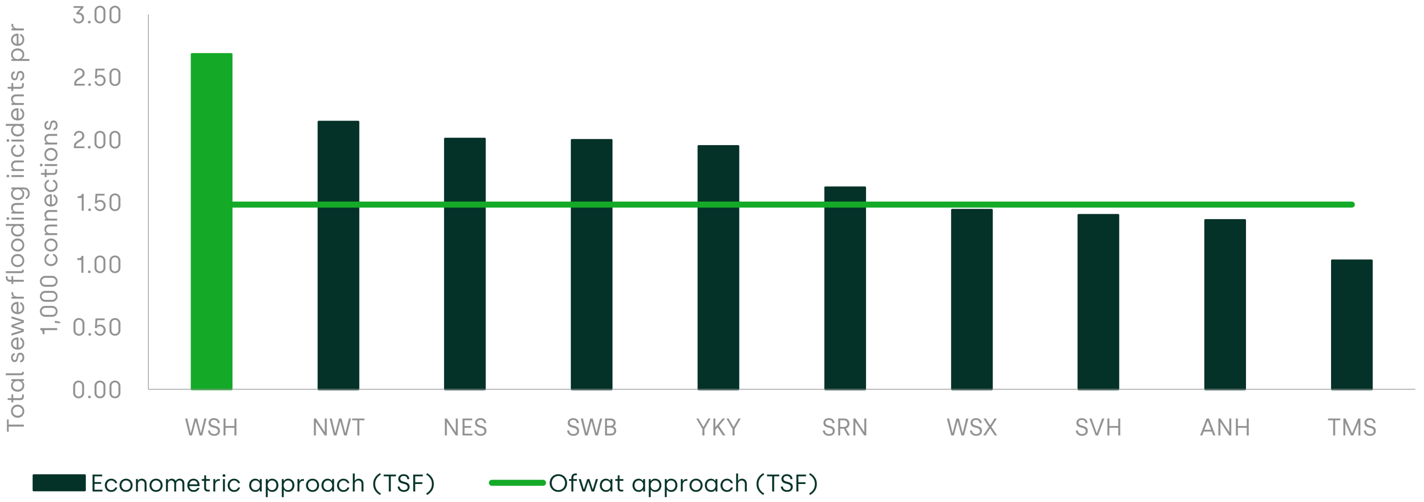
Source: Oxera analysis.

[8] The sewer flooding PCs (and therefore the dependent variable in the models) are defined per property, so the comparable models in the PR24 modelling consultation relate to some bioresources models and the residential retail models that are modelled on a unit-cost basis.
[9] At PR19 Ofwat set a forward-looking upper-quartile benchmark for ISF incidents. Forward-looking information is not currently available, so here we focus on the outturn performance. While Ofwat did not set ESF PCs on the basis of an upper-quartile benchmark (and did not set a specific target for TSF), it did use companies' performance against the upper quartile when considering whether a company was a 'good' performer. As such, we present the outturn upper-quartile performance as an approximation of Ofwat's approach for these PCs. We also present the targets under an upper-tercile benchmark which UK regulators have used where model quality is insufficiently strong to employ an upper-quartile benchmark.

The table shows that the non-econometric approach may overestimate the extent to which WSH can reduce sewer flooding incidents relative to the econometric approach. As the latter approach explicitly accounts for exogenous operational characteristics, this suggests that WSH operates in a particularly difficult region with respect to sewer flooding incidents. While the effect is directionally the same across ISF and ESF, the magnitude is significantly more material in ESF—i.e. WSH operates in a region where it is particularly difficult to reduce ESF incidents.

WSH is not the only company affected by the use of econometric models in setting PCs—the figure below shows how the industry's predicted performance under the econometric approach compares with that under the non-econometric approach.

Total sewer flooding performance (2019–23): incidents per 1,000 properties



Source: Oxera analysis.

The figure shows that the non-econometric approach—which does not account for relevant operational drivers of performance—could set more stringent targets for six out of the ten wastewater companies than the econometric approach. However, WSH is uniquely affected, as it operates in the most complex region according to these models.

Given that it is feasible to develop robust econometric models for sewer flooding in this way, Ofwat could explore such models when setting PCs at PR24. This could involve open engagement with the industry to further improve upon the models presented in this report and to ensure that all relevant drivers of various performance metrics are considered. We note that econometric analysis is one of many methods that can be used to normalise the PC targets, and could be reinforced by company-specific deep dives and further operational evidence.

1 Introduction

In a competitive market, companies are incentivised to provide a higher quality of service at an efficient cost. If a company in a competitive market provides a higher quality of service for a reasonable price, it can attract more demand and earn additional profits. Conversely, if a company provides a poorer quality service at a higher price, it will lose demand to other companies and earn lower profits (or incur larger losses). As natural monopolies, water companies are not subject to these competitive pressures, and (in the absence of regulation) there is a risk that consumers will suffer from poor service quality and/or high prices. Therefore, regulators typically set targets relating to both efficient costs (and therefore tariffs) and service quality (e.g. performance commitments, PCs).

At PR19, Ofwat used econometric modelling to set efficient cost allowances for the majority of companies' costs. In essence, companies' historical (and business plan) costs were benchmarked against those of the most cost-efficient companies in the industry, once a set of relevant exogenous factors were accounted for. Given the heterogeneity across the water industry across a range of characteristics (e.g. scale, population density and topography), regional factors need to be accounted for to ensure that comparisons are made between companies on a like-for-like basis. Meanwhile, the PCs were set using two different methods.

Common PCs were determined using comparative benchmarking models.[10] In contrast to Ofwat's approach on cost assessment, Ofwat did not account for exogenous characteristics when determining the appropriate PC. For example, the PC on leakage was determined by the volume of leakage per length of main of the upper-quartile company, while the PC relating to internal sewer flooding (ISF) was based on the number of ISF incidents per sewer length of the upper-quartile company. For some common PCs, the performance target was determined by a rate of expected improvement, rather than by achieving a specific level. Common PCs were typically applied to performance measures that had

[10] Common PCs include: water supply interruptions, internal sewer flooding, pollution incidents, leakage, per-capita consumption, water quality compliance, treatment work compliance, mains repairs, unplanned outages, sewer collapses, risk of severe restrictions in a drought, risk of sewer flooding in a storm, priority services register, customer measures of experience, and business measures of experience. See Ofwat (2019), 'PR19 final determinations: Delivering outcomes for customers policy appendix', December, section 3.

a (relatively) long time series of performance data, and where all companies were expected to make improvements. We understand that Ofwat has not provided an explicit justification for its approach, but has stated that it could see no ‘clear reasons why companies should not achieve the same stretching levels of performance’,[11] and that these metrics are 'both important to customers and supported by good quality comparative information’.

Bespoke PCs were determined using a bottom-up assessment of companies’ proposals.[12] Bespoke PCs were applied for some performance measures where there was insufficient historical data to undertake a comparative assessment, or where they affected only a few companies. As these PCs are bespoke and assessed using bottom-up methods, they can (in principle) account for company-specific characteristics.[13]

One limitation of Ofwat’s PR19 approach is that the PCs and cost targets were set independently, and there was minimal consideration of how companies’ efficient expenditure requirements would be affected by stretching PCs. By setting stretching, independent targets on both costs and service quality, there was a risk that the overall challenge facing companies would be unachievable. Indeed, this issue was raised by the appellant companies at the redetermination by the Competition and Markets Authority (CMA)—the CMA subsequently provided additional expenditure for companies to meet service targets.[14]

Another key limitation of Ofwat’s approach to setting common PCs is that Ofwat’s modelling does not account for relevant, exogenous company-specific characteristics. For example, it may be easier or harder for a company to achieve a certain level of ISF depending on the level of rainfall, how densely populated the region is, or the characteristics of the sewerage network. If these factors are unaccounted for when determining the PC, companies can make windfall gains or losses with respect to the ISF PC depending on their

[11] Ofwat (2019), ‘PR19 final determinations: Delivering outcomes for customers policy appendix’, December, p. 19.
[12] Bespoke PCs include: sewer blockages, external sewer flooding, low pressure, water quality customer contacts, residential gaps and voids, WINEP and NEP initiatives, and carbon and project-specific PCs. See Ofwat (2019), ‘PR19 final determinations: Delivering outcomes for customers policy appendix’, December, sections 3 and 6.
[13] For example, Ofwat set bespoke PCs for business customer satisfaction for Welsh companies but not for English companies because of the extent of competition in Wales. See Ofwat (2019), ‘PR19 final determinations: Delivering outcomes for customers policy appendix’, December, p. 155.
[14] See Competition and Markets Authority (2021), ‘Anglian Water Services Limited, Bristol Water plc, Northumbrian Water Limited and Yorkshire Water Services Limited price determinations Final report’, March, para. 4.495 and table 8.8.

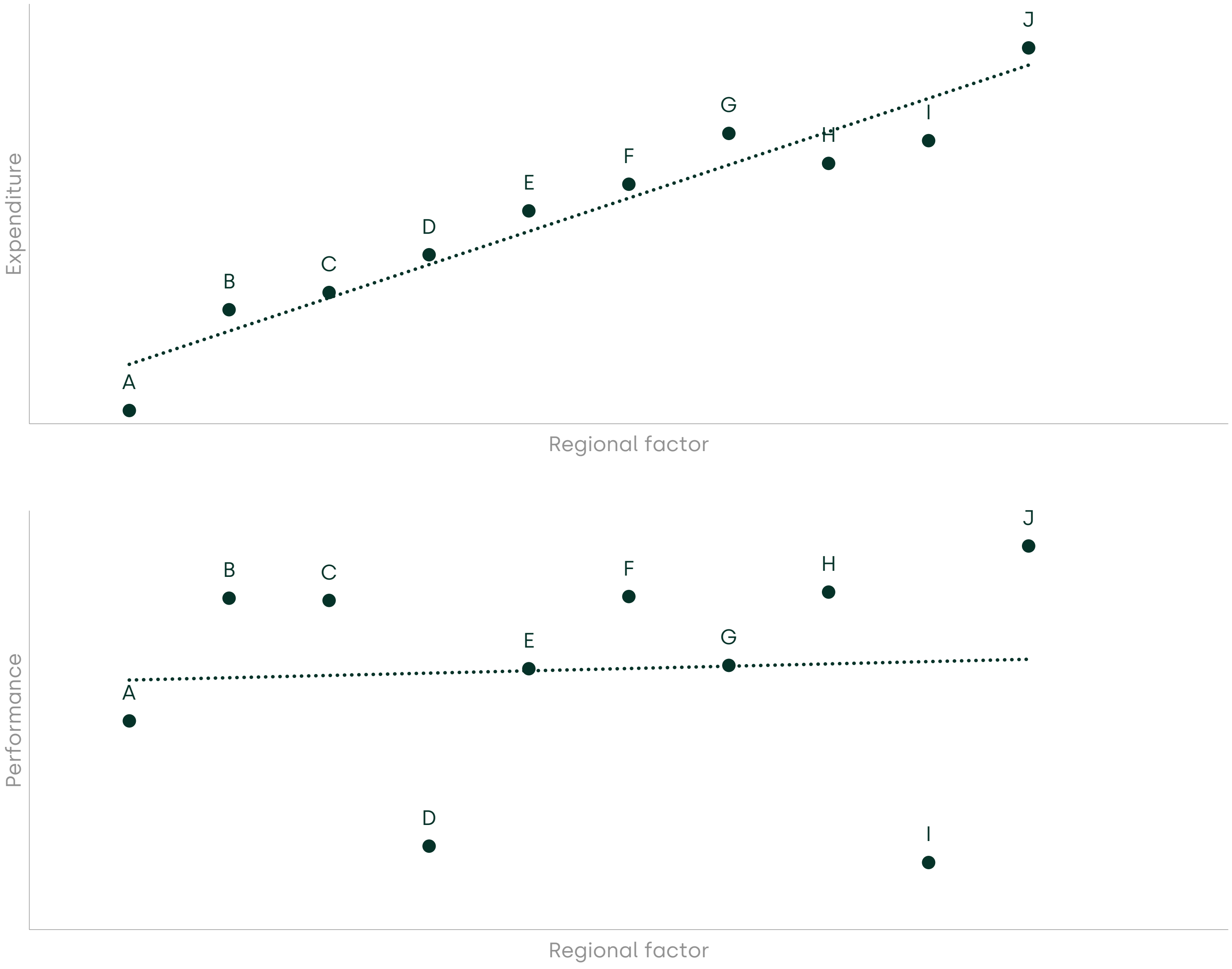
operating environment, rather than on how well they handle ISF incidents.

In its PR24 methodology, Ofwat stated that it would consider exogenous characteristics when determining companies' PC levels.[15] Accounting for exogenous characteristics that influence service quality is essential in order to demonstrate that the PCs are sufficiently challenging and achievable for all companies.

Ofwat further argued in its PR24 methodology that adjusting PCs on the basis of regional factors that are already accounted for in the base cost models may result in double-counting.[16] This argument rests on the assumption that the base models fund companies on the basis of achieving the same PC targets—an argument that can be tested empirically by assessing the correlation between companies' performance on these service measures and the regional factors. The figure below shows a stylised example of a case in which Ofwat is correct.

[15] Specifically, Ofwat stated: 'United Utilities, Northumbrian Water and Yorkshire Water expressed concerns that the measure does not consider exogenous factors impacting on the measure, such as the effect of high levels of rainfall on the operation of storm overflows, or the level of investment required to achieve the performance commitment […] we will consider evidence about relevant factors in setting expenditure allowances and/or performance commitment levels.' See Ofwat (2022), 'Creating tomorrow, together: Our final methodology for PR24 Appendix 7 – Performance commitments', December, p. 71.
[16] See Ofwat (2022), 'Creating tomorrow, together: Our final methodology for PR24 Appendix 9 – Setting expenditure allowances', December, p. 61.

Figure 1.1 Stylised example—models fund companies to achieve the same PCs

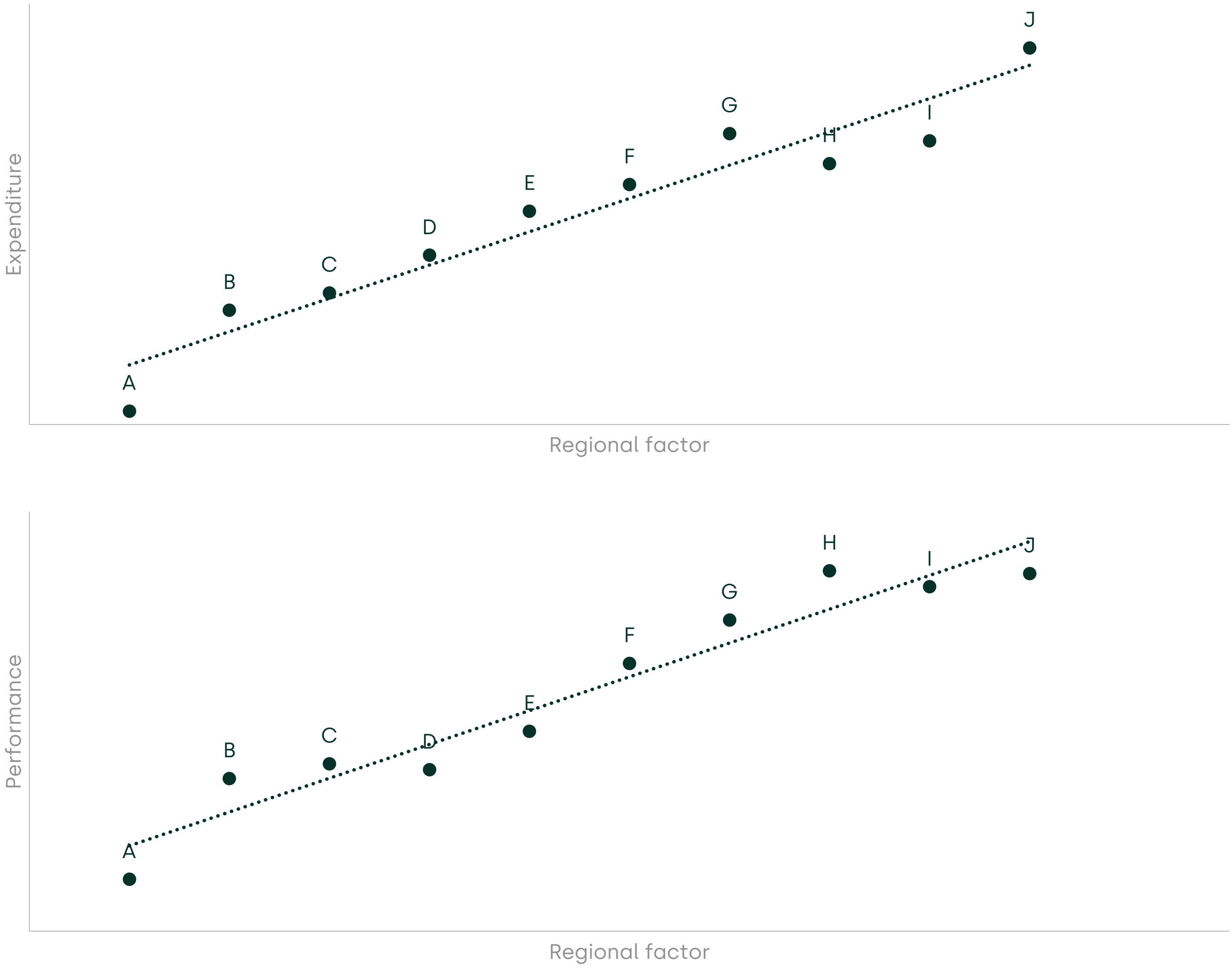


Note: Each point represents a company.
Source: Oxera.

In this stylised example, the regional factor included in the cost model has little correlation with the service performance. Here, it could be argued that the cost models fund companies on the basis of a common performance target—the regional factor drives costs but not performance, so companies should be able to deliver common PCs once the regional factor is reflected in the cost models.

The chart below shows a counterexample where Ofwat's assertion is incorrect.

Figure 1.2 Stylised example—models fund companies to achieve the same PCs



Note: Each point represents a company.
Source: Oxera.

In this example, there is a strong correlation between the regional factor and service performance. Because the regional factor drives both cost and performance, the inclusion of the regional factor in the cost models does not mean that companies are funded to deliver the same service level. For example, company A is funded on the basis of achieving a poorer performance than company J.

Therefore, Ofwat's concerns relating to the potential for a double-counting of regional factors when setting cost and performance targets can be empirically tested.

Indeed, this point is apparent in the way in which Ofwat currently normalises its PCs. For example, it sets the supply interruptions PC on the basis of the average supply interruption per property. Therefore, the number of connected properties (a regional factor) is already accounted for when setting PCs, even though four of the five base cost models already control for connected properties. Ofwat does not argue that, because its cost models already account for scale, scale does not need to be reflected when setting PCs. In the same way, simply because the cost models account for other regional factors (e.g. density) does not mitigate the need for the PCs also to account for other regional factors, providing that these regional factors are operationally relevant.

Dŵr Cymru (Welsh Water, WSH) has commissioned Oxera to identify key drivers of sewer flooding performance (both ISF and external sewer flooding, ESF), and to normalise for these drivers in an econometric framework consistent with Ofwat's approach to assessing efficient base costs. We also draw out implications from the econometric analysis on how these normalised performance targets ('econometric targets') compare with Ofwat's approach of benchmarking at the upper-quartile level without normalisation ('non-normalised targets').

This report is structured as follows.

- Section 2 outlines our methodology for developing sewer flooding models.
- Section 3 presents the models developed under the methodology outlined in section 2.
- Section 4 concludes.

2 Methodology

2.1 The dataset

Our modelling dataset is derived from three sources: (i) Ofwat's wastewater base cost modelling consultation dataset;[17] (ii) Ofwat's historical performance trends for PR24 dataset;[18] and (iii) an industry datashare for the 2022/23 annual performance reports (APRs) that we received from WSH. The data from the base cost modelling dataset and the 2022/23 APR datashare covers all companies from 2012 to 2023. Meanwhile, the data from the performance trends dataset (used to construct sewer flooding variables) and the APR datashare is from 2017 to 2023 for most of the industry, with the exception of Thames Water (TMS), which did not report ESF data in the 2022/23 APR datashare.

Following Ofwat's approach to modelling base expenditure in wholesale wastewater, for all of the analysis presented in this report we merge the data for Severn Trent England (SVE) and Hafren Dyfrdwy (HDD). HDD is a significantly smaller company than the rest of the sector, and is owned by the same entity as SVE. Therefore, merging the data for these companies mitigates the risk of HDD skewing the empirical analysis (this is the same logic as applied by Ofwat and the CMA).

Some external data used in Ofwat's base cost modelling dataset is not reported in companies' APRs.[19] For these variables, we have made the following assumptions to estimate the 2023 value.

- Urban rainfall excluding soil permeability (BON code BN4507): 2023 value determined based on the average value in 2018–22. Urban rainfall is volatile from year to year, so we consider that a smoothing approach is appropriate.
- Potential evapotranspiration (PET): 2023 value determined based on the average value in 2018–22. PET is volatile from year to year, so we consider that a smoothing approach is appropriate.
- Weighted average density—MSOA: 2023 value determined based on a linear extrapolation of historical data (2012–22). This

[17] Ofwat (2023), 'PR24 Cost Assessment Master Dataset, Wholesale Wastewater Base Costs v4', April, https://www.ofwat.gov.uk/wp-content/uploads/2023/04/PR24-Cost-Assessment-Master-Dataset-Wholesale-Wastewater-Base-Costs-v4.xlsx, accessed 11 August 2023.
[18] Ofwat (2023), 'Historical performance trends for PR24 V2.0', April, https://www.ofwat.gov.uk/publication/historical-performance-trends-for-pr24-v2-0/, accessed 9 August 2023.
[19] These variables are discussed in section 2.3.

approach captures the overall upward trend in population density.

- Proportion of population living in coastal areas: 2023 value set equal to the 2022 value. This assumes that the proportion of the population living in coastal regions is fairly constant.

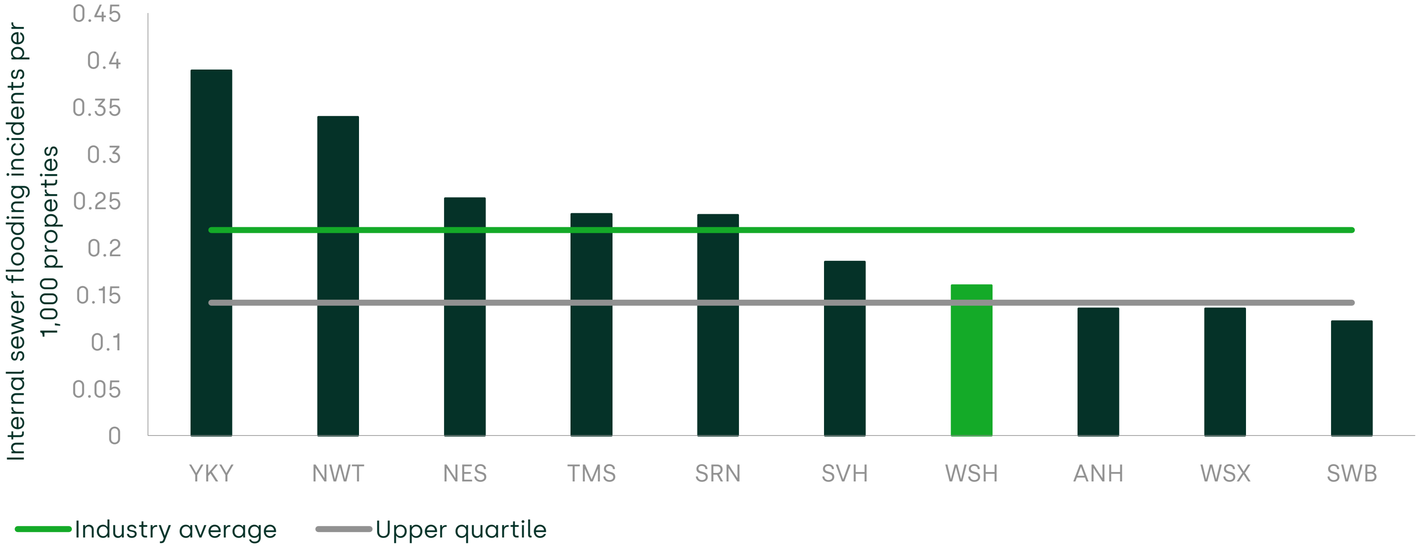
2.2 Defining the measures of service performance

A flooding incident is defined as the escape of water from the sewerage system. ISF is generally defined as when the flooding enters a main building, while ESF is defined as when the flooding occurs in the curtilage of the main building. If the flooding incident affects both the main building and the curtilage, it is reported as an ISF incident only. While Ofwat provides some guidance as to how companies should allocate flooding incidents between ISF and ESF,[20] there are incidents where a degree of judgement is required. For example, in the case of a flooding incident on a farm, companies are required to provide a 'reasonable allowance' for what would constitute a garden (in which case, it is in the curtilage of the building and is considered ESF) and what is not (in which case, the incident is outside of the curtilage and is not reported).

The chart below shows the distribution of ISF incidents per thousand properties across the industry (2019–23), and how companies perform relative to the industry average and the upper quartile. We assess performance on a smoothed basis as sewer flooding is, to some extent, a stochastic event that depends on (among other factors) weather, climate and asset condition. As such, a company's performance on sewer flooding in any individual year may be unrepresentative of how well it performs in reducing the number of sewer flooding incidents. We have selected a five-year average as this is consistent with how Ofwat benchmarked costs at PR19, and the length of an asset management period (AMP).

[20] A description of ISF and ESF can be found in Ofwat (2018), 'Reporting guidance – Sewer flooding', https://www.ofwat.gov.uk/wp-content/uploads/2018/03/Reporting-guidance-sewer-flooding.pdf, accessed 10 August 2023.

Figure 2.1 Internal sewer flooding incidents (2019–23)

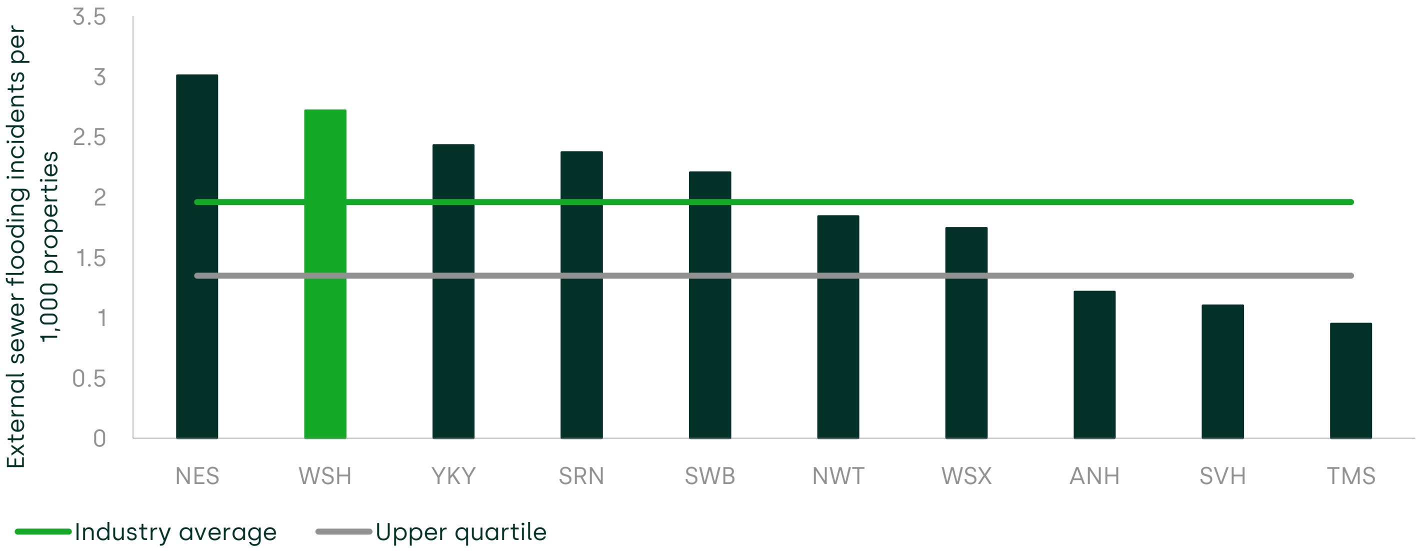


Source: Oxera analysis.

There is a large spread in performance on ISF across the industry. The worst-performing company on this measure has more than three times as many incidents per property than the best-performing company. Over the assessment period that we have considered, WSH has performed relatively well on this measure—it has significantly fewer ISF incidents than the industry average, and is fairly close to the upper-quartile performance.

The chart below shows the equivalent distribution for ESF incidents per thousand properties.

Figure 2.2 External sewer flooding incidents (2019–23)



Source: Oxera analysis.

As with ISF, the worst-performing company on this measure has more than three times as many incidents per property as the best-performing company. WSH performs worse on this metric than on ISF, and has more ESF incidents than the industry average.

WSH is not the only company whose performance materially differs between ESF and ISF. For example, SWB is the frontier company in relation to ISF, yet it performs worse than average on ESF. Meanwhile, TMS is the frontier company on ESF and performs worse than average on ISF. The inconsistency in companies' performances between ISF and ESF could point to any of the following issues:

- reporting inconsistencies—the distinction between ISF and ESF is not clear in all cases, and different companies may allocate similar flooding incidents differently;
- operational trade-offs—it is possible that steps taken to reduce ISF have knock-on effects on ESF, and vice versa;
- company focus—companies may focus attention (and funding) on improving different areas of ISF or ESF, depending on consumer preferences and the different incentives applied to ISF and ESF;
- operational characteristics—some operational characteristics (for example, population density) may have opposite effects on ISF and ESF.

To explore whether the inconsistent performance between ISF and ESF is an industry-wide issue or relates to isolated cases, the chart below shows the correlation between ESF and ISF across the industry.

Figure 2.3 Correlation between ISF and ESF (2019–23)

Source: Oxera analysis.

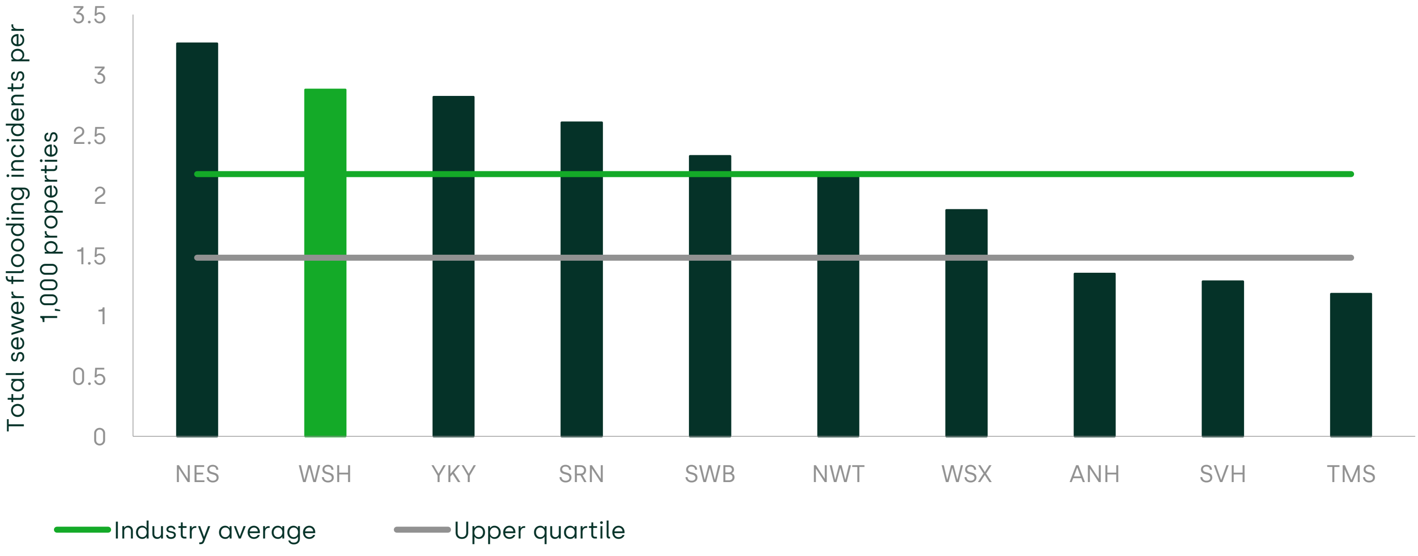
The figure shows that there is a weak positive correlation between ESF and ISF performance.[21] Given that the positive correlation is not strong, and there are several examples of companies that perform relatively well on one metric and relatively poorly on the other (e.g. SWB, TMS, WSH, NWT), the diverging performance between ISF and ESF remains an issue (even though the correlation is positive).

The observation that there is only a weak correlation between ISF and ESF performance, and that only one company has achieved an upper-quartile performance on both, could suggest that setting individual

[21] The correlation coefficient is 0.22.

upper-quartile targets on ESF and ISF separately could result in an unreasonably stringent benchmark. Therefore, we consider that it is appropriate to explore modelling ISF and ESF together (total sewer flooding, TSF). The distribution of TSF across the industry is shown in the figure below.

Figure 2.4 Total sewer flooding incidents (2019–23)



Source: Oxera analysis.

2.3 Performance drivers

We consider that sewer flooding—both ISF and ESF—could be driven by the following operational characteristics.[22]

- **Rainfall**. Inflows caused by rainfall are likely to put a strain on the sewerage network, and excessive rainfall could lead to sewer flooding. There are several measures of rainfall in the base modelling consultation dataset, including total rainfall and different measures of urban rainfall.
- **Population density**. The urbanity of operating areas may affect the propensity of different regions to flood. For example, ESF incidents occur in the curtilages of buildings, and are more likely to occur in regions with a higher volume of gardens, farms and

[22] There is overlap between the performance drivers outlined in this section and those considered in United Utilities' report for the future ideas lab on a similar topic. See United Utilities (2022), 'What lessons can we learn from cost assessment at PR19?', section 2.5.

golf courses, all of which are associated with lower population densities. Meanwhile, ISF incidents may occur more frequently in urban regions, where flooding due to poor drainage may be more common.

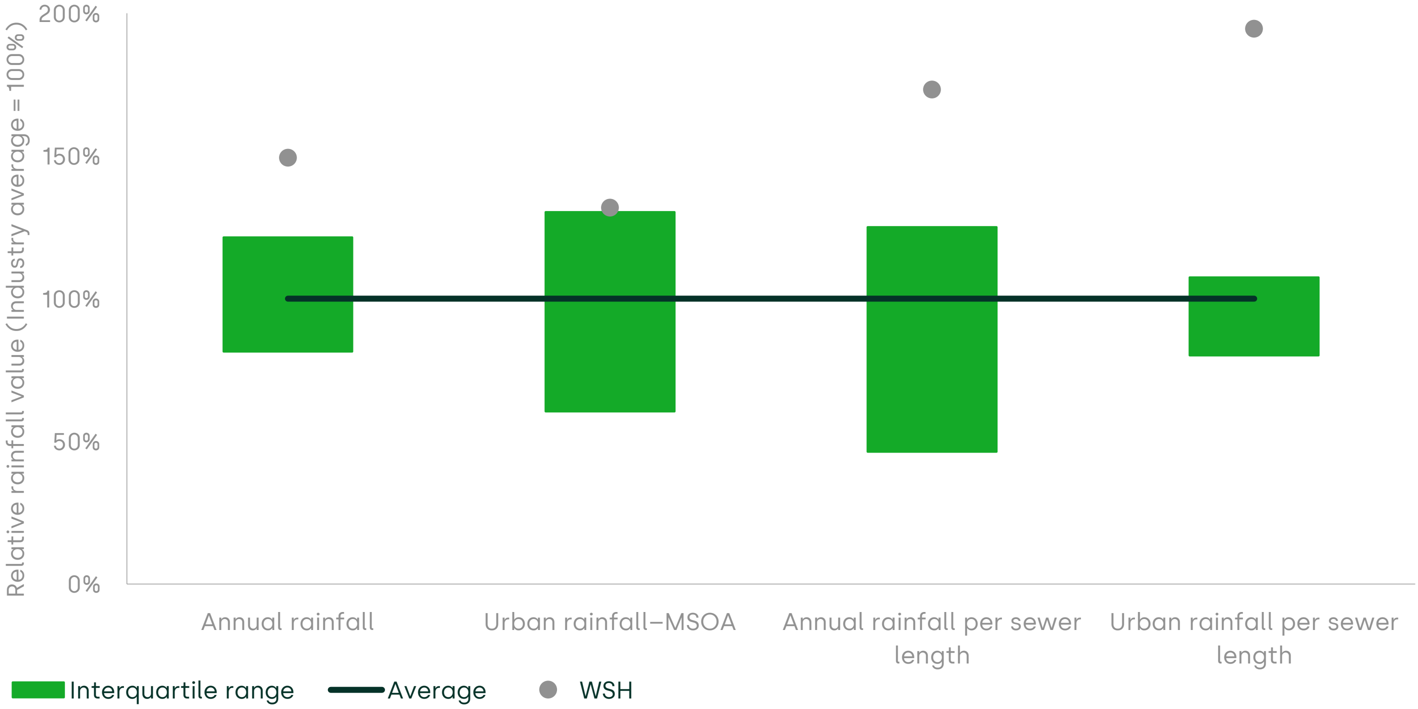
- **Potential evapotranspiration (PET)**. PET is a measure of the amount of evaporation that would occur in a region if there were an unlimited supply of surface water, and it is driven by (among other factors) temperature and wind speed. In principle, a region with high PET may be less prone to flooding as surface water evaporates more quickly than an area of low PET.[23] One would therefore expect there to be a negative relationship between PET and sewer flooding.
- **Network characteristics**. Certain asset types may be more prone to flooding than others. For example, as noted in some companies' responses to the PR24 cost modelling consultation, combined sewers are more prone to flooding than other sewer types because they operate near capacity for longer.[24]
- **Coastal population**. Sewers that connect to the coast are more susceptible to sudden rises in sea levels than those that do not. Higher sea levels as a result of storms or climate change can lead to backups, resulting in increased pressure on the sewerage system, which can increase the likelihood of flooding. There is also the increased risk of saltwater intrusion into sewerage systems, damaging assets and disrupting flow.
- **Atypical company performance**. Some companies may have atypical performance that cannot be explained by the cost drivers outlined above. These drivers might include historical enhancement expenditure or other operational characteristics (e.g. extreme weather or type of property served). For example, at PR19 YKY submitted a cost adjustment claim (CAC) to account for the high volume of cellared properties in its region. YKY argued that cellared properties are more prone to ISF incidents, thus inhibiting its ability to meet the same performance on ISF as the rest of the industry.[25] While Ofwat rejected this CAC at PR19 (on the basis that the CAC was quantified using outdated analysis), YKY does have significantly

[23] In theory, this effect could also be partially captured by the population density drivers. However, the correlation between the two drivers is fairly weak (c. 0.22). Therefore, it is unlikely that these drivers are capturing the same characteristics.
[24] See, for example, United Utilities (2023), 'UUW response - Consultation on econometric base cost models for PR24', April, p. 19.
[25] See Ofwat (2019), 'Cost adjustment claim feeder model Yorkshire Water', December, https://www.ofwat.gov.uk/wp-content/uploads/2019/12/FM_CAC_YKY_FD_.xlsx, accessed 10 August 2023.

more ISF incidents per property than the rest of the industry, and this YKY-specific characteristic may need to be appropriately accounted for when developing robust ISF models.

The charts below show how these performance drivers are distributed across the industry.

Figure 2.5 Distribution of rainfall drivers across the industry (2019–23)



Note: MSOA refers to Middle Super Output Area.
Source: Oxera analysis.

Across the various rainfall measures, WSH's operating region is consistently wetter than that of the upper quartile of the industry. Furthermore, WSH encounters the second-highest volume of rain per sewer length in general, and the highest volume of urban rainfall per sewer length. This indicates that WSH faces significant pressure on its sewer network relating to rainfall, which could result in increased sewer flooding.

The table below shows the correlation between these rainfall metrics and sewer flooding performance.

Table 2.1 Correlations between performance and rainfall metrics

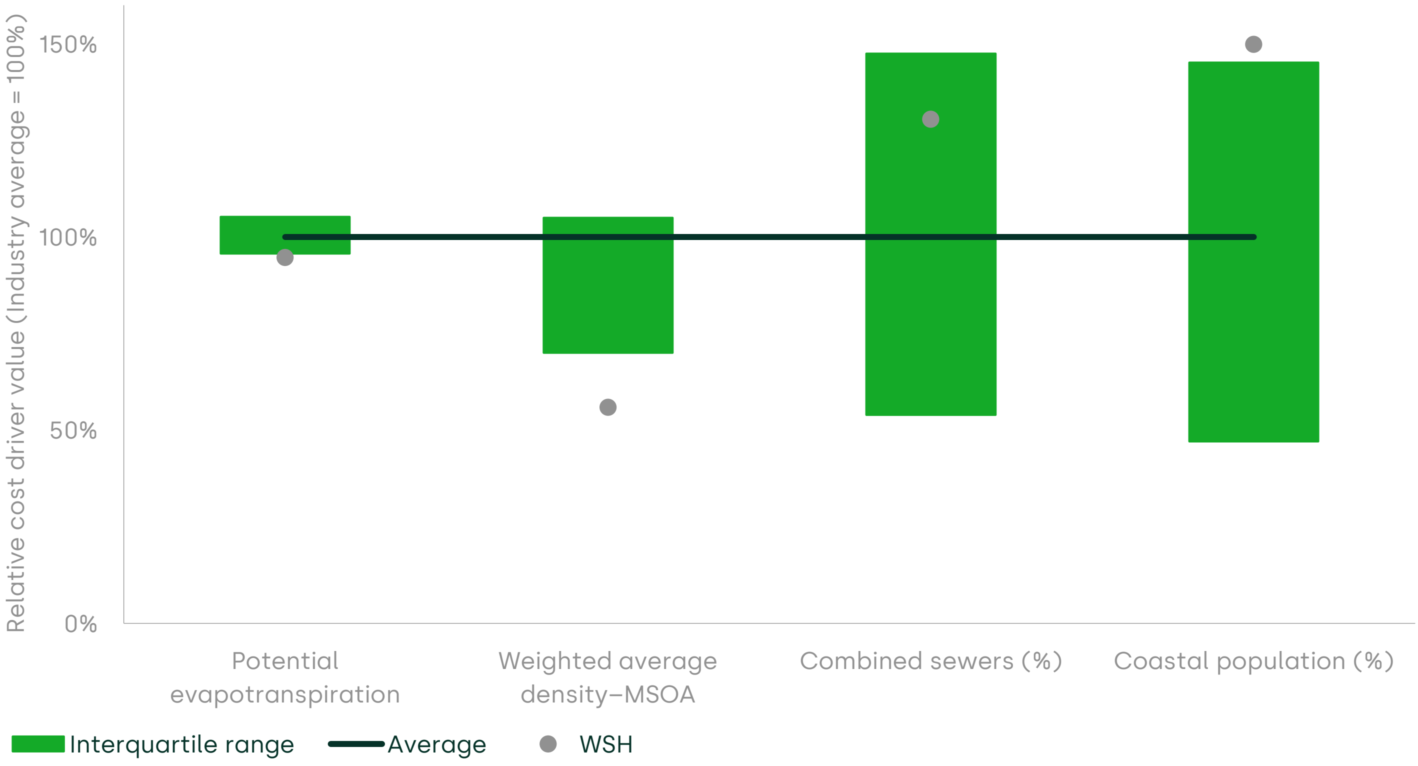
	ISF	ESF	TSF
Annual rainfall	0.0639	0.4929***	0.4849***
Urban rainfall—MSOA	0.3962***	-0.2672**	-0.1981
Annual rainfall per sewer length	-0.1489	0.7541***	0.6984***
Urban rainfall per sewer length	0.2054*	0.5443***	0.5449***

Note: * p<0.1, ** p<0.05, *** p<0.01. All variables are in logs.
Source: Oxera analysis.

The table shows that most rainfall metrics are positively correlated with companies' sewer flooding performance. However, annual rainfall does not have a strong correlation with ISF (indeed, it is negative when normalised per sewer length), while urban rainfall is negatively correlated with ESF and TSF when it is not normalised by sewer length. Urban rainfall per sewer length is the only rainfall metric that has a positive and statistically significant correlation with all measures of sewer flooding. It may be the case that the rainfall metric that is most appropriate differs by the type of sewer flooding incident.

The figure below shows the equivalent analysis for the other performance drivers.

Figure 2.6 Distribution of other performance drivers across the industry (2019–23)



Source: Oxera analysis.

Across the potential performance drivers, WSH appears to be unique in terms of population density, as its population is the most rural under this measure. WSH is also below the lower quartile in terms of PET, and above the upper quartile for the proportion of its population residing on the coast. Combining these factors with the aforementioned operational rationale for driving performance, it appears ex ante that

WSH operates in a region that is particularly susceptible to sewer flooding.

The table below shows the correlation between these other performance drivers and sewer flooding performance.

Table 2.2 Correlations between performance and other drivers

	ISF	ESF	SF
Potential evapotranspiration	-0.3745***	-0.5356***	-0.5637***
Weighted average density—MSOA	0.3324***	-0.4720***	-0.3893***
Combined sewers (%)	0.3766***	0.6007***	0.6227***
Coastal population (%)	-0.2518**	0.6320***	0.5650***

Note: * p<0.1, ** p<0.05, *** p<0.01. All variables are in logs unless otherwise stated in parentheses.
Source: Oxera analysis.

All correlations are statistically significant at standard thresholds. PET is negatively correlated with sewer flooding across the different performance measures, while combined sewers is positively correlated with all sewer flooding measures. Population density is positively correlated with ISF and negatively correlated with ESF, indicating that companies that operate in densely populated regions typically perform worse on ISF and better on ESF (and TSF). These correlations are broadly aligned with the operational expectations outlined above.

Coastal population is strongly positively correlated with ESF and TSF, which is also aligned with operational expectations. However, it is negatively correlated with ISF, and the operational argument for this is less clear. There is a significant correlation between density and coastal populations (-0.5080)—in the case of ISF, the correlation with coastal populations may be spurious since density is strongly correlated with ISF.

We note that the correlations presented in this section are 'univariate'—i.e. they point towards performance drivers that may perform best in the models in isolation; however, alternative performance drivers may be superior when modelled alongside other performance drivers.

2.4 Model development process

The aim of the econometric modelling is to capture as many operationally relevant drivers of performance as necessary, while ensuring that the models are not over-specified ('parsimonious') and robust from an operational, economic and statistical perspective. This can be broken down as follows.

- **Operational performance**. The operational quality of the model is based on an assessment of whether all the relevant drivers of performance have been considered in the assessment, and whether the estimated relationship between sewer flooding and performance drivers is aligned with operational intuition. In principle, this would assess whether both the *direction* and the *magnitude* of the estimated relationship are aligned with expectations. However, given the fairly short history of modelling PCs (relative to cost modelling), the expected magnitude of the relationship between sewer flooding and performance drivers would require further research. We note that, even in Ofwat's cost models, more work is needed to assess whether the magnitude of the estimated relationship between costs and cost drivers is operationally appropriate.[26] Our operational assessment focuses on whether the direction of the relationship is aligned with expectations.
- **Economic performance**. The economic quality of the model relates primarily to whether the models could induce or reflect perverse incentives. For example, if there were a negative correlation relationship between asset health and sewer flooding (i.e. healthier assets had fewer flooding incidents), this could encourage companies to let their assets deteriorate to achieve more relaxed PCs (or reward companies that have let their assets deteriorate).
- **Statistical performance**. The statistical quality of a model is determined through the statistical significance of the coefficients on the performance drivers, the degree of model fit, and the performance of other statistical diagnostic tests (such as the RESET test for model specification). We note that—as in Ofwat's cost modelling—most statistical tests are built on

[26] This was also noted by Professor Andrew Smith (Ofwat's academic adviser) in the PR24 cost modelling consultation. See Ofwat (2023), 'Econometric base cost models for PR24', April, appendix A5.

assumptions that are unlikely to hold in the current context.[27] Therefore, the statistical performance of the models will be assessed qualitatively in the model development process.[28]

These criteria are aligned with Ofwat's criteria for developing econometric cost models.[29]

2.5 Benchmark selection

The econometric models predict companies' expected performance based on the average characteristics of the factors considered. At PR19, Ofwat set several PCs based on the upper-quartile performance rather than the average.

We note that the selection of the benchmark in the performance modelling has a direct parallel with the selection of the benchmark in the cost models. At PR19, Ofwat set the cost benchmark in wholesale wastewater at the third-ranked company. However, in the PR19 redetermination the CMA relaxed the stringency of the benchmark to the upper quartile, arguing (among other things) that the quality of the models did not support Ofwat's stringent challenge.[30]

The selection of the benchmark should be driven by the quality of the econometric models,[31] which should include an assessment of the degree of statistical uncertainty in the models. The focus of this report is on the development of econometric models rather than the selection of an appropriate benchmark. Therefore, in this report, we do not undertake a detailed assessment of the appropriate benchmark

[27] For example, the RESET test relies on the assumption that the residuals are independent and identically distributed and follow a normal distribution. If there is some level of inefficiency in the sample, or there is an 'optimal' level of performance that companies cannot surpass with a given technology, the residual will be one-sided; see Oxera (2021), 'A critical assessment of TCB18 electricity', April, Box 4.1. Moreover, some statistical tests are only asymptotically valid (i.e. only give close to the 'correct' answer in large samples) and are therefore inaccurate in small samples.
[28] This is consistent with Ofwat's treatment of statistical diagnostic tests. For example, see Ofwat (2023), 'Econometric base cost models for PR24', April, p. 16.
[29] Specifically, points 2 and 5 in Ofwat's principles of cost assessment. See Ofwat (2023), 'Econometric base cost models for PR24', April, figure 2.1.
[30] For example, see Competition and Markets Authority (2021), 'Anglian Water Services Limited, Bristol Water plc, Northumbrian Water Limited and Yorkshire Water Services Limited price determinations Final report', March, para. 4.493.
[31] We note that a careful consideration of the appropriate benchmark is not limited to econometric models. Under Ofwat's existing framework, which does not use econometric modelling to set PCs, there should also be an assessment of the quality of the comparisons in service performance measures. In this respect, we note that the quality of the comparisons under the existing framework is limited—Ofwat's framework does not account for several relevant exogenous drivers when determining PCs, and there can be a wide range of performance across the industry (equivalent to a wide range in estimated efficiency scores in the cost modelling). Therefore, even if Ofwat does not explore econometric modelling for PCs, it should carefully assess whether an upper-quartile benchmark is appropriate.

associated with these models. Instead, we present companies' performance relative to three benchmarks that were applied at PR19.

- Average benchmark—this was the benchmark that Ofwat applied to some cost models at PR19, and has been used in cases where the regulator has less faith in the econometric models (e.g. because costs are estimated with a high degree of uncertainty).
- Upper-quartile benchmark—this was the benchmark that Ofwat used to assess common PCs at PR19, and therefore provides a useful comparison to Ofwat's PR19 approach. However, such a benchmark implies that the regulator has a high degree of confidence in its models.
- Upper-tercile (i.e. upper-third) benchmark—this benchmark lies between the upper-quartile and average benchmarks.

3 Econometric models

This section presents the models developed under the criteria outlined in section 2.

3.1 Performance modelling of internal sewer flooding

We have developed six ISF models. In all cases, the dependent variable is the natural logarithm of the number of ISF incidents per property.[32] The models control for the following characteristics:

- urban rainfall per sewer length (all models);
- weighted average density (all models);
- PET (ISF2 and ISF4);
- combined sewers (ISF3 and ISF6);
- a dummy variable to reflect YKY-specific characteristics (ISF3–ISF6).[33]

We have also sought to control for other characteristics, such as soil permeability and coastal regions. However, these performance drivers performed relatively poorly in the model specifications (e.g. they had unintuitive and/or insignificant coefficients) on the current dataset. These drivers may need to be revisited in when developing ISF models for the PR24 determination when more (and/or improved) data becomes available.

The table below shows the proposed models for ISF per property.

[32] This is consistent with how Ofwat sets the performance targets with respect to ISF (and ESF) incidents.
[33] As outlined in section 2.3, YKY has argued that it faces more difficulty in reducing ISF incidents due to the prevalence of cellared properties in its region. A YKY dummy can be considered an (albeit blunt) instrument to capture this effect, given that data regarding the number of cellared properties is not readily available.

Table 3.1 ISF modelling

	ISF1	ISF2	ISF3	ISF4	ISF5	ISF6
Urban rainfall per sewer length (log)	0.504***	0.301*	0.237*	0.428***	0.230*	0.260*
	(0.00204)	(0.0553)	(0.0886)	(0.00275)	(0.0839)	(0.0530)
Weighted average density—MSOA (log)	0.604***	0.656***	0.876***	0.572***	0.623***	0.790***
	(0.000145)	(1.14e-05)	(7.32e-09)	(4.41e-05)	(1.28e-06)	(8.68e-08)
Potential evapotranspiration (log)		-2.746***			-2.693***	
		(0.000432)			(5.67e-05)	
Proportion of combined sewers (%)			3.166***			2.380***
			(2.54e-07)			(0.000229)
YKY dummy variable				0.736***	0.727***	0.420**
				(1.17e-05)	(1.69e-06)	(0.0119)
Constant	-4.838***	11.80**	-8.365***	-4.880***	11.44***	-7.513***
	(4.79e-05)	(0.0126)	(1.44e-10)	(3.86e-06)	(0.00447)	(4.18e-09)
Observations	70	70	70	70	70	70
Adjusted R-squared	0.206	0.333	0.463	0.399	0.525	0.505
RESET	0.546	0.754	0.197	0.433	0.345	0.343
VIF	1.141	1.307	1.426	1.157	1.321	1.897

Note: The dependent variable is the natural logarithm of the number of ISF incidents per 1,000 properties. The model is estimated over the period 2017–23. P-values are given in parentheses. * p<0.1, ** p<0.05, *** p<0.01.
Source: Oxera analysis.

The table shows that there is a positive and statistically significant relationship between urban rainfall and ISF across a range of model specifications, as is the case with the relationship between population density and ISF. The coefficients on PET and proportion of combined sewers are statistically significant and of intuitive signs. The coefficient on the YKY dummy variable is positive and statistically significant, which indicates that YKY has systematically higher levels of ISF per property once the other cost drivers are accounted for, and is a potential outlier requiring appropriate treatment.

The models 'pass' the RESET test and there is fairly low multicollinearity between the performance drivers (as indicated by the low VIF). While the model fit is lower than that for most of Ofwat's base cost models, note that the dependent variable is measured in terms of ISF per property, and that scale (a material driver of total ISF incidents) is already captured in the construction of the independent variable and is not reflected in the R-squared. Compared with Ofwat's proposed unit cost models (in 'other retail costs' and bioresources), these ISF models have a higher model fit (by c. 0.1–0.4).

The table below shows how WSH performs in these models on an outturn basis (2019–23), compared with the non-econometric approach.

Table 3.2 ISF models predicted performance levels for WSH

	Non-econometric approach	ISF 1	ISF2	ISF3	ISF 4	ISF5	ISF 6	Triangulated
Outturn performance	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
Predicted performance (average benchmark)	0.21	0.23	0.21	0.19	0.20	0.19	0.19	0.20
Predicted performance (upper-tercile benchmark)	0.16	0.18	0.18	0.17	0.17	0.17	0.16	0.17
Predicted performance (upper-quartile benchmark)	0.14	0.17	0.17	0.16	0.17	0.16	0.16	0.16
Rank	4	2	1	2	2	3	2	2

Source: Oxera analysis.

WSH performs relatively well in these models—it is always ranked within the upper quartile (top three out of ten) and is ranked second on a triangulated basis. This is better than WSH's performance according to Ofwat's analysis (which looks simply at companies' performance based on ISF per property), where WSH ranks fourth. On a triangulated basis, WSH's performance is c. 21% better than the average benchmark and c. 1% better than the upper quartile. That is, WSH recorded 21% fewer incidents than expected given its unique operating circumstances.

3.2 Performance modelling of external sewer flooding

We have developed five ESF models. In all cases, the dependent variable is the natural logarithm of the number of ESF incidents per property. The models control for the following characteristics:

- urban rainfall per sewer length (all models);
- weighted average density (ESF1–ESF4);
- PET (ESF2, ESF4 and ESF5);
- combined sewers (ESF3–ESF5);

- proportion of coastal regions (ESF5).

The table below shows the models for ESF per property.

Table 3.3 ESF modelling

	ESF1	ESF2	ESF3	ESF4	ESF5
Urban rainfall per sewer length (log)	0.521***	0.365***	0.376***	0.333***	0.154*
	(6.13e-05)	(0.00323)	(0.00239)	(0.00667)	(0.0863)
Weighted average density—MSOA (log)	-0.389***	-0.354***	-0.243**	-0.277**	
	(0.00199)	(0.00221)	(0.0430)	(0.0208)	
Potential evapotranspiration (log)		-2.109***		-1.337*	-1.000*
		(0.000487)		(0.0631)	(0.0511)
Proportion of combined sewers (%)			1.730***	1.064*	1.850***
			(0.000561)	(0.0739)	(1.51e-05)
Proportion of population that reside on the coast (%)					1.682***
					(0)
Constant	5.244***	18.07***	3.343***	12.20**	6.854**
	(1.91e-07)	(4.04e-06)	(0.00112)	(0.0131)	(0.0392)
Observations	69	69	69	69	69
Adjusted R-squared	0.373	0.473	0.471	0.491	0.736
RESET	0.0939	0.152	0.132	0.200	0.00832
VIF	1.123	1.286	1.387	2.175	1.907

Note: ESF data for TMS in 2023 is not available, so this observation has been removed from the sample. The dependent variable is the natural logarithm of the number of ESF incidents per 1,000 properties. The model is estimated over the period 2017–23. P-values are given in parentheses. * p<0.1, ** p<0.05, *** p<0.01.
Source: Oxera analysis.

The table shows that urban rainfall is positive and statistically significant across specifications, while population density is negative and significant. The coefficients on PET, combined sewers and proportion of coastal population are all statistically significant and (directionally) aligned with operational expectations.

The adjusted R-squared is typically higher in the ESF models than in the ISF models. Most models 'pass' the RESET test, with the exception of ESF5. This could indicate that higher-order terms (such as squared terms and interactions) are omitted from the model. We have explored the inclusion of higher-order terms in the model, but the implied elasticities are inconsistent with operational expectations for some observations. Therefore, we do not consider that failing the RESET test is a material concern in ESF5.

The table below shows WSH's performance in these models.

Table 3.4 ESF models predicted performance levels for WSH

	Non-econometric approach	ESF 1	ESF2	ESF3	ESF 4	ESF 5	Triangulated
Outturn performance	2.72	2.72	2.72	2.72	2.72	2.72	2.72
Predicted performance (average benchmark)	2.02	3.36	3.15	3.08	3.06	2.80	3.09
Predicted performance (upper-tercile benchmark)	1.74	2.72	2.54	2.33	2.41	2.64	2.45
Predicted performance (upper-quartile benchmark)	1.34	2.48	2.48	2.30	2.35	2.39	2.44
Rank	9	4	5	5	5	6	5

Source: Oxera analysis.

WSH performs better than the average benchmark in all models and on a triangulated basis. On a triangulated basis, it has c. 12% fewer ESF incidents than the average benchmark and c. 9% more incidents than the upper-quartile benchmark. Note that WSH's performance in these models is significantly better than under Ofwat's simple approach, where WSH ranks nine out of ten.

3.3 Performance modelling of total sewer flooding

We have developed five TSF models. In all cases, the dependent variable is the natural logarithm of the number of TSF incidents per property. These models control for similar cost drivers to the ESF models, as follows:

- urban rainfall per sewer length (all models);
- weighted average density (TSF1–TSF3);
- PET (TSF2–TSF4);

- combined sewers (TSF3–TSF4);
- proportion of coastal regions (TSF4).

The table below shows the results from the TSF models.

Table 3.5 TSF modelling

	TSF1	TSF2	TSF3	TSF4
Urban rainfall per sewer length (log)	0.523***	0.360***	0.321***	0.157*
	(3.66e-05)	(0.00240)	(0.00542)	(0.0831)
Weighted average density—MSOA (log)	-0.261**	-0.224**	-0.133	
	(0.0297)	(0.0384)	(0.232)	
Potential evapotranspiration (log)		-2.207***	-1.285*	-1.097**
		(0.000152)	(0.0578)	(0.0348)
Proportion of combined sewers (%)			1.270**	1.742***
			(0.0244)	(4.75e-05)
Proportion of population that reside on the coast (%)				1.367***
				(5.47e-09)
Constant	4.359***	17.78***	10.78**	7.686**
	(4.86e-06)	(2.27e-06)	(0.0196)	(0.0226)
Observations	69	69	69	69
Adjusted R-squared	0.326	0.452	0.486	0.692
RESET	0.0684	0.132	0.0995	0.00370
VIF	1.123	1.286	2.175	1.907

Note: The dependent variable is the natural logarithm of the number of TSF incidents per 1,000 properties. The model is estimated over the period 2017–23. P-values are given in parentheses. * p<0.1, ** p<0.05, *** p<0.01.
Source: Oxera analysis.

The TSF models listed in Table 3.5 perform at a level similar to the ESF models in terms of model fit. Urban rainfall is statistically significant with a positive magnitude, while PET and weighted average density are statistically significant and negative.[34] The proportion of combined sewers is positive and statistically significant, which is aligned with operational expectations.

The table below presents WSH's performance across the TSF models.

Table 3.6 TSF models predicted performance levels for WSH

	Non-econometric approach	TSF 1	TSF2	TSF3	TSF 4	Triangulated
Outturn performance	2.88	2.88	2.88	2.88	2.88	2.88
Predicted performance (average benchmark)	2.25	3.57	3.34	3.22	3.07	3.30
Predicted performance (upper-tercile benchmark)	1.88	2.88	2.75	2.63	2.67	2.72
Predicted performance (upper-quartile benchmark)	1.48	2.64	2.71	2.50	2.67	2.68
Rank	9	4	5	5	6	5

Source: Oxera analysis.

WSH's performance in the TSF models is comparable to its performance in the ESF models—it performs better than the average benchmark across all models and on a triangulated basis, but has a gap to the upper quartile. On a triangulated basis, WSH has had c. 13% fewer incidents than the average benchmark and c. 6% more incidents than the upper-quartile benchmark. Again, even though there is a gap to the upper quartile in these models, the gap is significantly narrower than under Ofwat's simple approach that does not account for exogenous factors.

3.4 Triangulation

In this report, we present several models at different levels of performance that meet our model selection criteria. At this stage, we consider that the models are equally informative according to our criteria: the bottom-up models (modelling ESF and ISF separately) could better reflect the specific drivers affecting ISF and ESF (e.g. the observation that density is positively associated with ISF and negatively

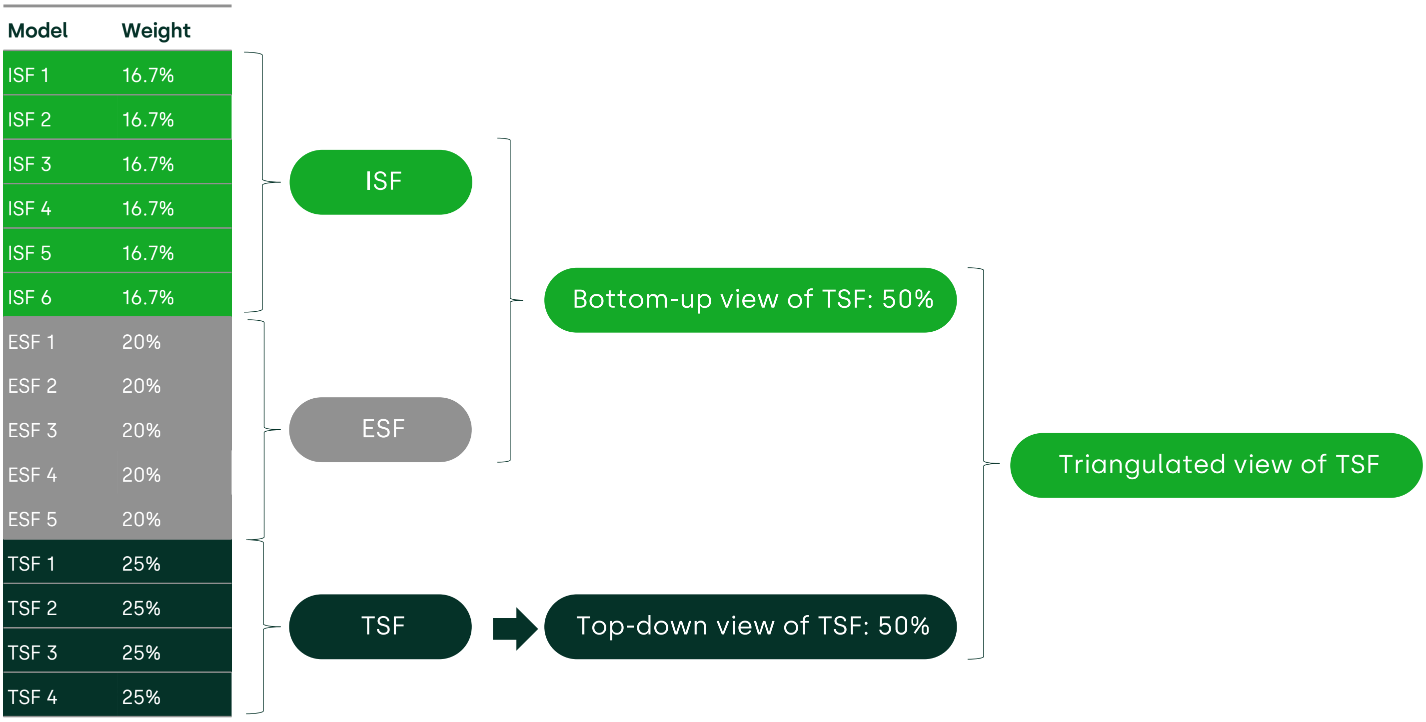
[34] Albeit with the exception of weighted average density in TSF3. It is important to note that this significance level is similar to that of the weighted average treatment complexity variable in the water resources plus models in the proposed PR24 cost models.

associated with ESF), while the top-down models (modelling TSF) could better account for the operational trade-offs between ISF and ESF (outlined in section 2.2).

An overall assessment of companies' sewer flooding performance could be conducted through a triangulation approach. In this report, we adopt a simple average approach to triangulating the results across models, broadly following Ofwat's approach to triangulating its wholesale base models at PR19, where it could not discriminate between models on the basis of reliability or confidence in the models.

The triangulation approach is illustrated in the figure below.

Figure 3.1 Triangulation approach



Source: Oxera.

Such a triangulation approach provides a point estimate for companies' overall sewer flooding targets. However, Ofwat may still wish to set separate targets on ISF and ESF (e.g. because consumers may have different preferences over companies' ISF performance versus their ESF performance). These separate targets can be determined based on an individual company's relative performance on ISF and ESF.[35] For example, c. 5.6% of WSH's TSF relates to ISF, so its ISF PC is determined as c. 5.6% of the triangulated view of TSF.

The table below shows how WSH performs on a triangulated basis.

Table 3.7 Triangulated performance levels for WSH

	ISF	ESF	Bottom-up	TSF	Triangulated
Outturn performance	0.16	2.72	2.88	2.88	2.88
Predicted performance (average benchmark)	0.20	3.09	3.29	3.30	3.30
Predicted performance (upper-tercile benchmark)	0.17	2.45	2.74	2.72	2.73
Predicted performance (upper-quartile benchmark)	0.16	2.44	2.68	2.68	2.68
Rank	2	5	5	5	5

Source: Oxera analysis.

WSH's performance does not differ materially between the bottom-up and top-down models. WSH has incurred c. 13% fewer flooding incidents than the average benchmark and c. 6% more than the upper-quartile benchmark.

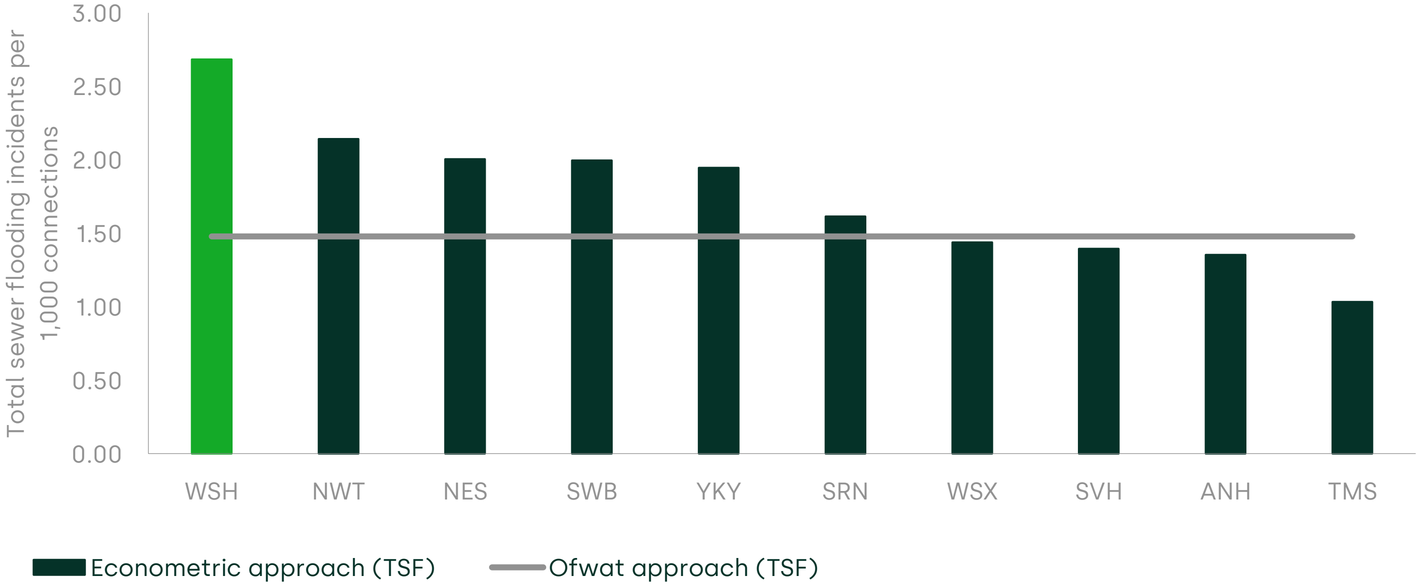
[35] This approach of pro-rating the performance has its limitations. For example, an individual company may find it easier to achieve a more stringent target in ISF than in ESF. An alternative approach would be to use the bottom-up models to disentangle the triangulated TSF prediction. Nonetheless, we consider that a pro-rating approach is proportionate at this stage, and it broadly aligns with Ofwat's approach of setting allowances in the wholesale price controls.

4 Concluding remarks

The analysis presented in the preceding sections demonstrates that it is feasible to develop econometric models for assessing sewer flooding performance. The statistical robustness of these models is broadly comparable to that of the cost models that Ofwat presented as part of the PR24 cost modelling consultation. Moreover, the use of econometric models mitigates the risk that Ofwat's approach to setting PCs is biased in favour or against specific companies as a result of exogenous operational characteristics.

The figure below demonstrates this argument by comparing companies' predicted TSF in the econometric models (at the upper-quartile benchmark) with the upper-quartile TSF in the industry (the latter being an approximation of Ofwat's approach to PCs).[36]

Figure 4.1 TSF upper-quartile performance distribution: econometric versus Ofwat approach



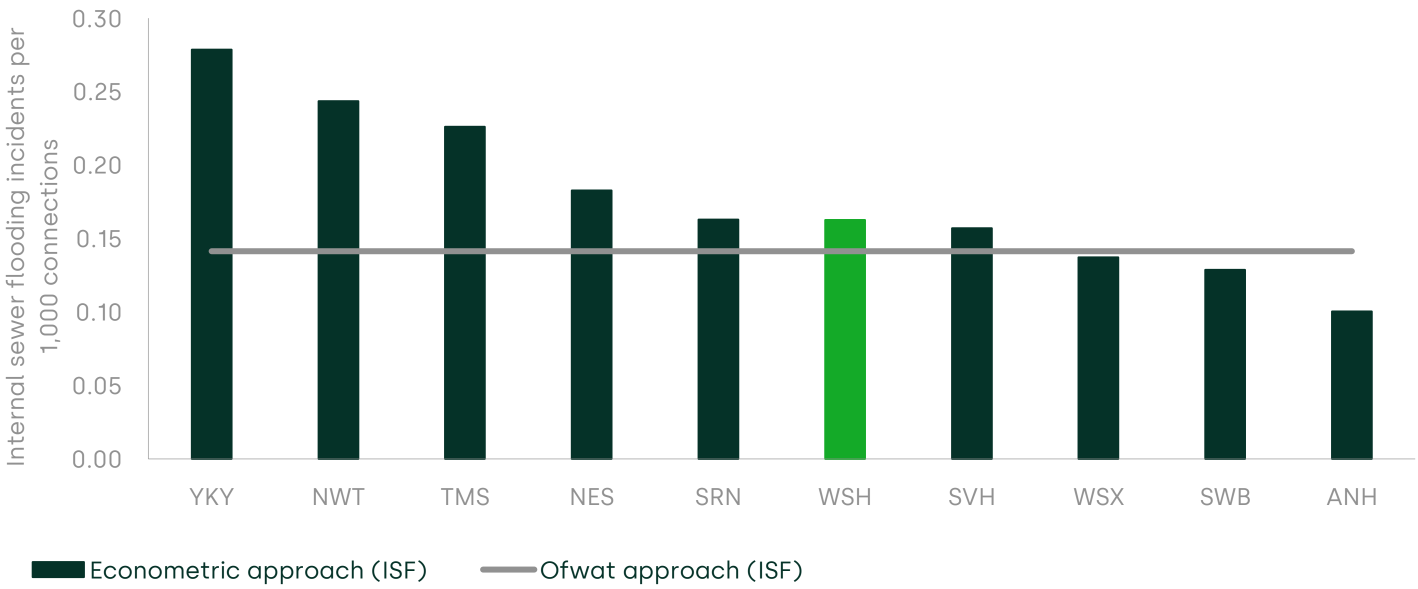
Source: Oxera analysis.

[36] Ofwat did not set a single target for TSF; rather, it set an ISF target on the basis of a forward-looking upper quartile and an ESF target based on a more bespoke analysis.

The figure shows that the upper quartile-corrected prediction from the econometric models is significantly lower than the upper-quartile service performance for some companies such as TMS. According to the models, these companies have operating characteristics that make it easier to achieve a set level of TSF. Meanwhile, the figure shows that several companies, in particular WSH, operate in regions that are particularly complex (the upper quartile-corrected predicted performance is substantially above the upper-quartile performance level). For these companies, the upper-quartile performance level is significantly more challenging.

The following figures extend the aforementioned analysis to the individual PCs, starting with the ISF models.

Figure 4.2 ISF upper-quartile performance distribution: econometric versus Ofwat approach

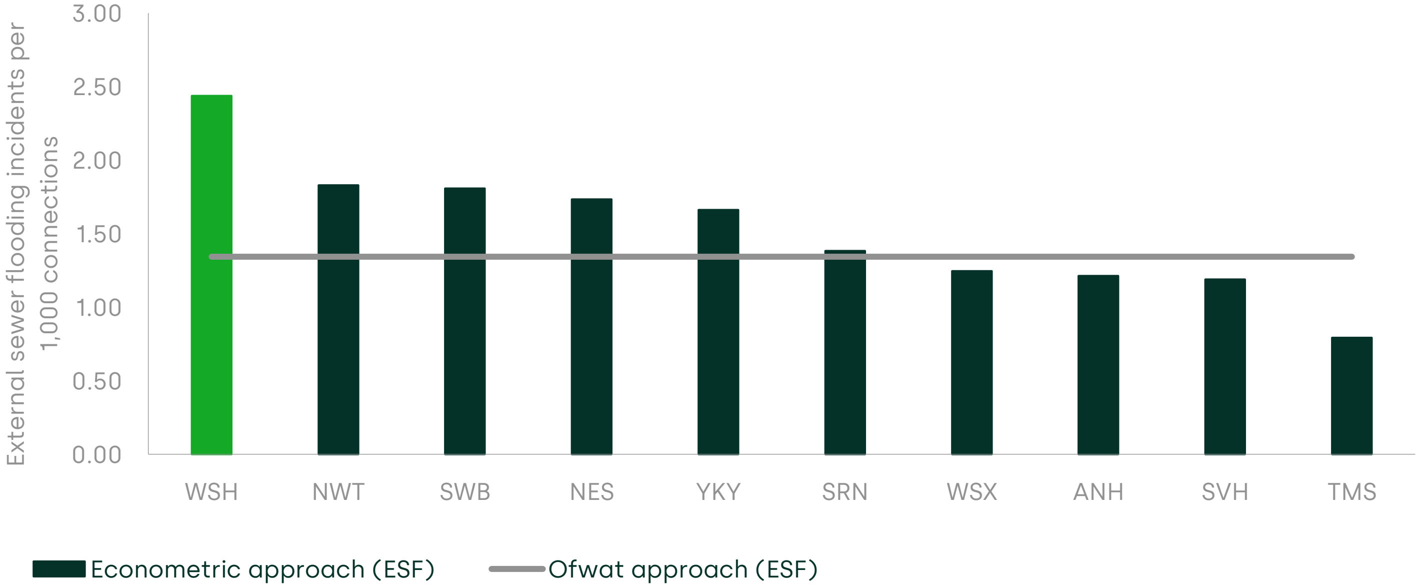


Source: Oxera analysis.

Figure 4.2 shows that there is a significant disparity across companies' abilities to meet ISF targets. The figure shows that the upper quartile-corrected performance based on the econometric models for WSH is higher than the upper-quartile service performance across the industry, indicating that Ofwat's approach (of taking a simple upper quartile on the performance measure) overestimates WSH's ability to improve its ISF performance.

The figure below shows the equivalent analysis for ESF.

Figure 4.3 ESF upper-quartile performance distribution: econometric vs Ofwat approach



Source: Oxera analysis.

Figure 4.3 is broadly similar to Figure 4.1[37] (on TSF modelling)—WSH has operating characteristics that make it significantly more difficult to perform at the industry upper-quartile service level. That is, the expected number of ESF incidents for WSH, having accounted for drivers of performance, is substantially higher than the outturn industry upper-quartile level. This demonstrates the significant disadvantage that WSH would face if Ofwat continued to set targets using the existing approach.

Given that it is feasible to develop robust econometric models for sewer flooding in this way, Ofwat could explore such models when setting PCs at PR24. This could involve open engagement with the industry to further improve upon the models presented in this report and to ensure that all relevant drivers of performance are considered. We note that econometric analysis is one of many methods that can be used to

[37] This is expected, given that ESF forms the majority of TSF across the industry (c. 90%).

normalise the PC targets, and could be reinforced by company-specific deep dives and further operational evidence.

Contact

Dr Srini Parthasarathy
Partner
+44 (0) 20 7776 6612
srini.parthasarathy@oxera.com

oxera.com