Response to econometric model consultation

United Utilities Water

04 May 2018

Contents

1 Overview

We are pleased to have the opportunity to review and comment on the econometric cost models that are being considered, both by Ofwat and by other companies, in advance of PR19. The price control process overall is strengthened by this increase in transparency, which facilitates informed consideration of the issues associated with cost assessment.

We are supportive of Ofwat's overall approach to modelling, as set out in its consultation:

- Use engineering, operational and economic understanding to specify an econometric model, and form expectations about the relationship between cost and cost drivers in the model.
- Assess whether the estimated coefficients are of the right sign and of plausible magnitude.
- Consider if the estimated coefficients are robust. For example, are they stable and consistent across different specifications? Are the estimated coefficients statistically significant?
- Assess the consequences of cost drivers under management controls, in particular, the risk of any perverse incentive.
- Consider the statistical validity of the model more widely – does the model perform well in terms of statistical tests and diagnostics?
- Consider the appropriate estimation method.

Many of our comments on the models that were shared on 29 March 2018 relate to one or more of the above principles. Whilst we are also providing detailed feedback on individual models, our review of the various models has been particularly helpful in surfacing a number of higher level themes, which we explore further in this document. This also provides some additional colour to the above principles and how they should be applied in practice. Some key themes which we explore below are as follows:

- **Sensitivity to model results to individual companies or groups of companies (cf. criteria 3)**
- **Models should avoid variables where there is little variation (or where one company provides the majority of industry variation) (cf. criteria 3)**
- **Models should contain factors which represent the broad range of service delivery, not overly focus on one aspect (cf. criteria 1)**
- **Models should avoid explanatory variables correlate with scale, even where this is only relatively weak (cf. criteria 1 and 2)**
- **Prior assumptions need to be tested with companies (cf. criteria 2)**

In addition to this, we have also provided in this document our main overall findings from the models representing each of the main value chains. We would be more than happy to discuss any findings set out in this response in more detail with Ofwat (the following section also summarises our preferred models from those provided from the consultation).

There are two things that we would like to draw particular attention to, on the water models:

- Many of the proposed water models appear overly weighted towards pumping and network related cost drivers, and do not attempt to capture key differences in treatment related drivers and (in particular) the costs of managing and treating different sources. For us, this makes a large proportion of the water models unacceptable for use in an industry benchmarking model.
- Furthermore, many of the proposed aggregated models only include an incomplete component of pumping head. We are very concerned and strongly believe that each of the measures should reflect the corresponding pumping head value across the relevant services, consistent with the aggregation of cost within the model. There also appears to be some material differences in reporting methodology for that cost driver which accentuates the problems – this is particularly concerning

when as it often has a very high leverage within the models, seeming to predict larger variations in cost than it should.

Given the very large number of models to review in such a short timescale, we will no doubt continue to keep working through them in advance of submitting our main PR19 business plan later this year. We will provide any further insights into this modelling work along with our business plans.

1.1. United Utilities' preferred models from the consultation

Water	Wastewater	Retail
Wholesale Water	**Wholesale Wastewater**	**Retail Total Cost**
WSHWW1	YKYWWW3	ORTC3
BRLWW2	OWWW4	ORTC4
SEWWW4	UUWWW1	UURTC1
		UURTC2
Water Network plus	**Wastewater Network plus**	
BRLNPW2	ONPWW3	**Retail Bad Debt Cost**
SVTNPW1	UUNPWW1	ORDC1
		ORDC2
Treated Water Distribution	**Sewage Treatment**	
OTWD1	OSWT5	**Retail Remaining Cost**
WSXTWD1	UUSWT1	OROC1
		OROC2
Water Resources plus	**Sewage Collection**	UUROC1
No suitable models	UUSWC1	UUROC2
	OSWC2	
Water Resources		
OWR1	**Bioresources plus**	
SRNWR3	OBP4	
	OBP5	
	OBP6	
	OBP7	
	Bioresources	
	WSHBR1	
	UUBR2	
	WSXBR1	
	WSXBR4	

The models listed above are those which we believe are best suited to construct the basic cost threshold within cost assessment for PR19. One important point is that we believe there is not one model that we consider the 'perfect' model; indeed the likelihood is that such a model will not be possible given the constraints in the data available. We have selected a suite of models from those presented in the consultation across each of the appropriate value chains, which constitute a selection of models that complement one another and through effective triangulation, offer the best chance for a robust basic cost threshold to be generated against which the benchmarking can begin. We have not yet proposed any model(s) for the Water Resources plus aggregation as we believe that there still improvements that can be made in order to achieve the most suitable suite of models.

The following tables demonstrate how the combination of models and aggregations fit together in order to generate basic cost thresholds for each of the price controls.

Table 1 - Water

Value chain	Top down	Bottom Up	Resources plus
Water Resources		✓	
Water Resources plus			✓
Treated Water Distribution			✓
Water Network plus		✓	
Water	✓		
Water treatment	n/a	n/a	n/a
Wholesale Water (plus)	n/a	n/a	n/a

Table 2 - Wastewater

Value chain	Top down	Bottom Up	Resources plus
Bioresources		✓	
Bioresources Plus			✓
Sewage Collection		✓	✓
Sewage Treatment		✓	
Network Plus Wastewater		✓	
Wholesale Wastewater	✓		
Wholesale Wastewater (plus)	n/a	n/a	n/a

Table 3 - Retail

Value chain	Top down	Bottom Up
Bad Debt		✓
Remaining Cost		✓
Total Cost	✓	

2 Key summary points – issues with constructing water industry cost assessment models

In this section we set out some high level themes that have arisen in our reviews of the various models. The themes provide additional detailed considerations to Ofwat's six criteria, and should be given careful consideration when concluding on the final set of models to be utilised at PR19.

2.1 Sensitivity to model results to individual companies or groups of companies (cf. criteria 3)

Individual companies (or small numbers of companies) may have characteristics such that they exert significant influence of on the magnitude and, potentially, the signage of the co-efficient(s) within a model. Ofwat should treat such models with extreme care, to ensure that the unusual characteristics of one or more companies do not inadvertently result in illegitimate cost predictions for other companies.

For example, evidence from the econometric models suggests Portsmouth Water is a highly influential company in cost assessment exercises. Portsmouth has high "leverage" to model outcomes, suggesting that it has more weight in the fit of econometric models than the average company. Importantly, this high leverage stems from Portsmouth's unusual profile across all variables used in Ofwat's models – not simply the company's size. Portsmouth's high leverage has significant influence on model results, which is highlighted by the impact of excluding the company on coefficients across wholesale water models. Despite Portsmouth representing only 0.7% of industry botex, its exclusion leads to a 3% redistribution of modelled costs across the industry and has a significant impact on efficiency benchmark assessment. This is problematic as changes in coefficients and the redistribution of modelled costs reflects omitted variable biases.

Overall, we believe that outlier analysis and company exclusion robustness tests can improve the quality of models utilised at PR19. This can then better inform the extent to which those model assessments (and assessment of efficiency benchmarks) can and should be relied upon when estimating the efficient costs of individual companies. High sensitivity to the inclusion of individual companies may suggest that a variable is behaving as an endogenous company-specific factor or that the model is simply unstable. These criteria can be used to reject variables or models. Predicted cost baselines and efficiency challenges which explicitly account for robustness concerns are more likely to improve customer and company outcomes at PR19.

High leverage (on model predictions) may signify high statistical value, but not in combination with omitted variables. A company with extreme driver values can provide more information about the relationship between a driver and cost than companies clustered around the sample average. However, this does not hold true when there is a structural break in the dataset, or where including the company in question increases omitted variable biases. If companies with high leverage are also significantly affected by omitted variable biases, model coefficients will be attenuated and measures of relative efficiency will be misleading.

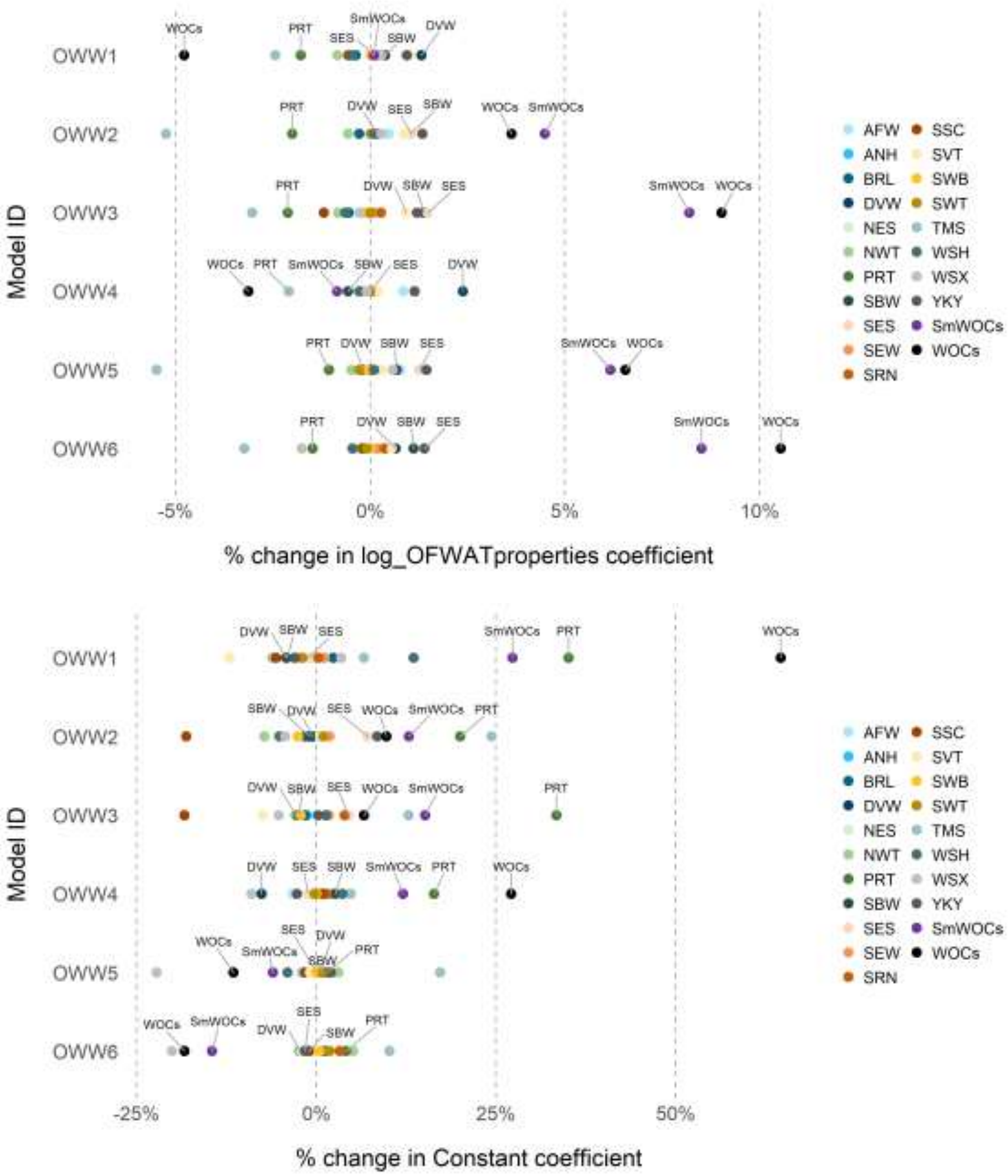
This illustrates one of the key risks arising from the small number of water company observations. Often one is reliant on a very small number of companies having characteristics which vary from the average in order to generate an apparently meaningful cost relationship. However, when that is due to one company - or a small number of companies – then there is a very high risk that the relationship identified is biased. This bias can be very significant, given the variation in scale between the smallest WoC and the largest WaSC.

The following series of charts illustrate the degree of variation in coefficients arising from the exclusion of individual companies (or groups, e.g. WoCs, or the group of smallest WoCs).

The exclusion of Portsmouth Water and small WOCs has a substantial impact on the connected properties coefficient in Ofwat's wholesale water models. Figure 1 shows the coefficients on connected properties, a key scale variable, and the constant change in Ofwat's wholesale water models, OWW1 – OWW6, when each company is dropped from the sample. Changes are small in percentage terms, but are highly material given the economic significance of connected properties in explaining wholesale water service costs. The exclusion of Portsmouth lowers the coefficient by around 2-3%, but with an offsetting rise in the constant term which varies in magnitude across the models.

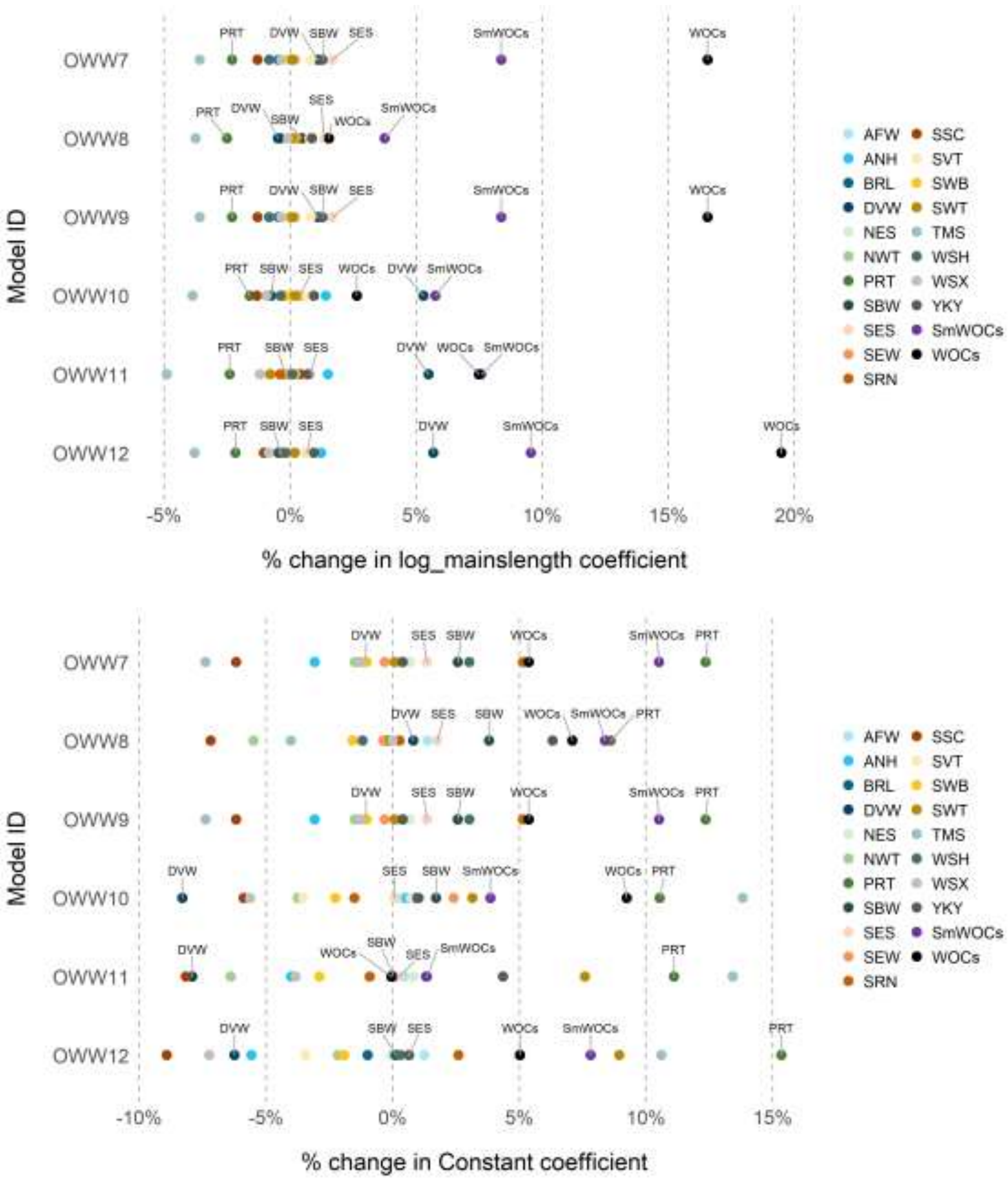
Results from models based on mains length are similar, albeit slightly larger in magnitude. This is shown in Figure 2. The coefficient on mains length in OWW7 – OWW12 is more sensitive to company exclusion than that for properties in OWW1 – OWW6. The mains length coefficient for models OWW10 – OWW12 is very sensitive to the inclusion of Dee Valley Water, the smallest WOC, with a rise of over 5% in each model when Dee Valley is removed from the sample. The exclusion of Portsmouth continues to lead to a fall in the scale variable coefficient, and a rise in the constant term.

Figure 1. **Changes in properties and constant term in Ofwat models OWW1 – OWW6 when companies are removed**



Note: *% change in coefficients for each Ofwat wholesale water model using properties as the key scale variable, as described in Cost assessment for PR19: Appendix 1 when each company is removed from the sample; WOCs – water only companies (AFW, BRL, DVW, PRT, SBW, SES, SEW, SSC); SmWOCs – small water only companies (DVW, PRT, SBW, SES), defined as WOCs serving <500,000 properties.*

Figure 2. **Changes in mains length and constant term in Ofwat models OWW7 – OWW12 when companies are removed**



Other model variables are less robust to excluding companies, with large swings in % mains relined, pumping head, asset age, and source type variables in wholesale and subservice models. This evidence can be used to identify and rule out variables which act as company dummies (i.e. where one company provides the majority of variation in a particular factor – see next section) and those which are simply not robust based on available data. The results are shown in the figures below as follows:

- Figure 3: Ofwat wholesale water model 1, OWW1, wherein % mains relined and % mains length post-1981 are highly unstable to excluding companies;

- Figure 4: Ofwat water network plus model 1, ONPW1, wherein % mains relined, average pumping head and booster pumping stations per km of mains length are highly unstable to excluding companies;
- Figure 5: Ofwat water resources plus model 1, OWRP1, for which source type and resources economies of scale findings change considerably when excluding Portsmouth Water.

Figure 3. **% mains relined and % mains length post-1981 are particularly unstable in Ofwat wholesale water model 1**

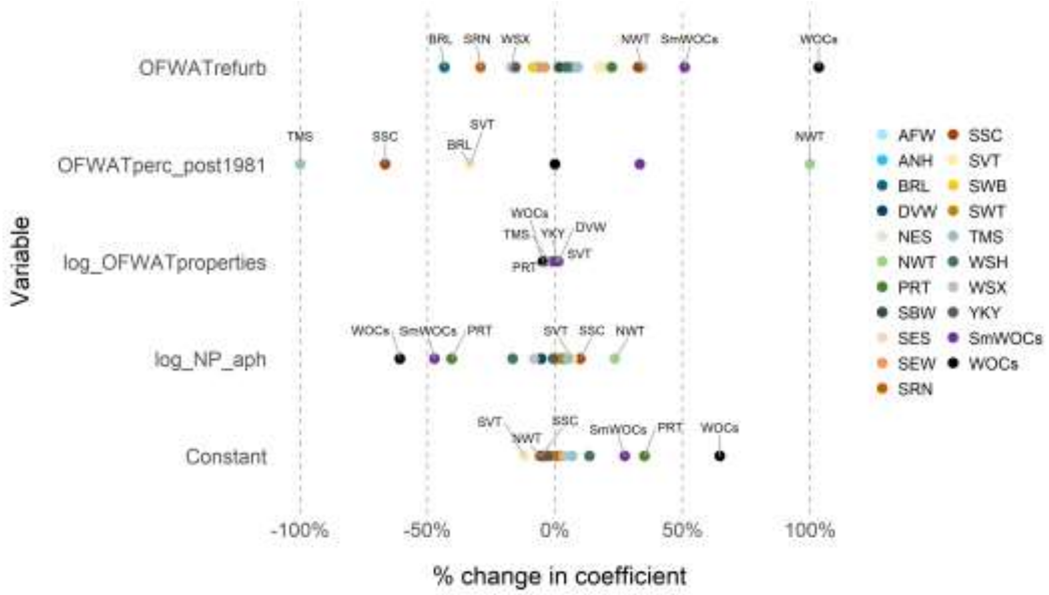


Figure 4. **Ofwat water network plus model 1 findings around % mains relined, average pumping head and booster pumping stations seem least stable**

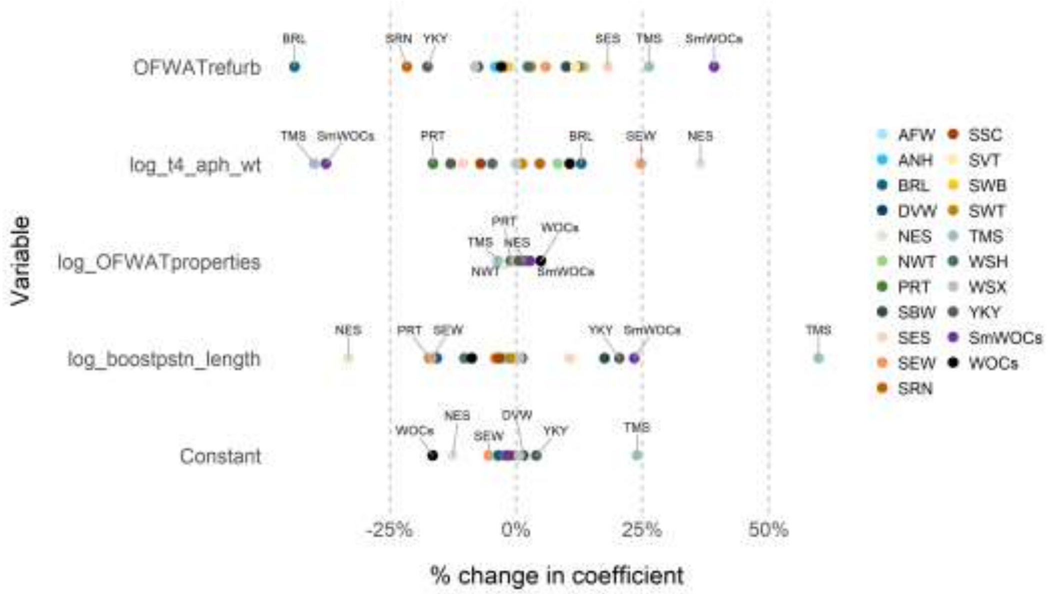
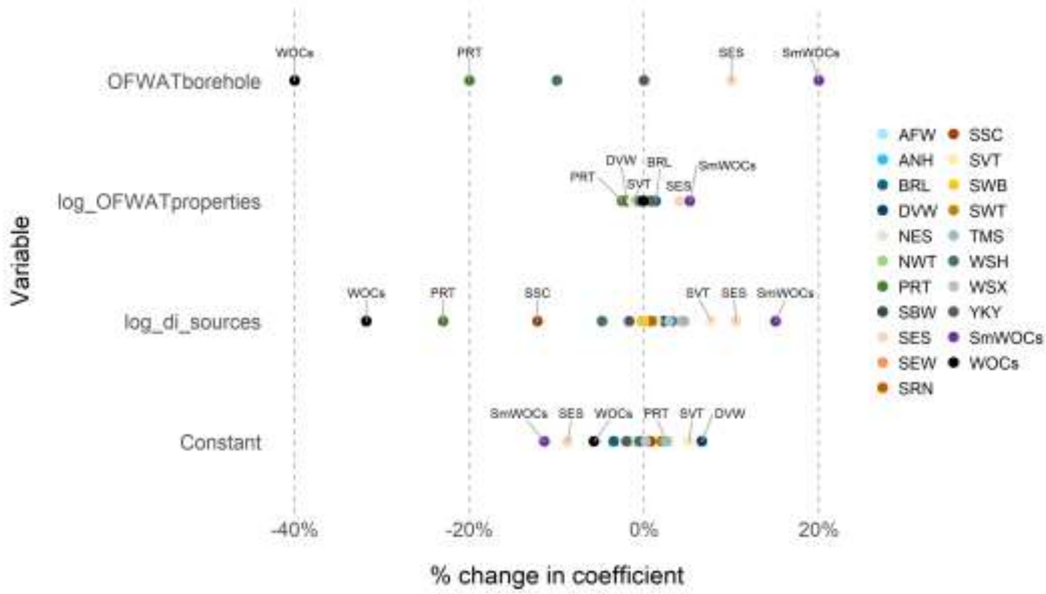
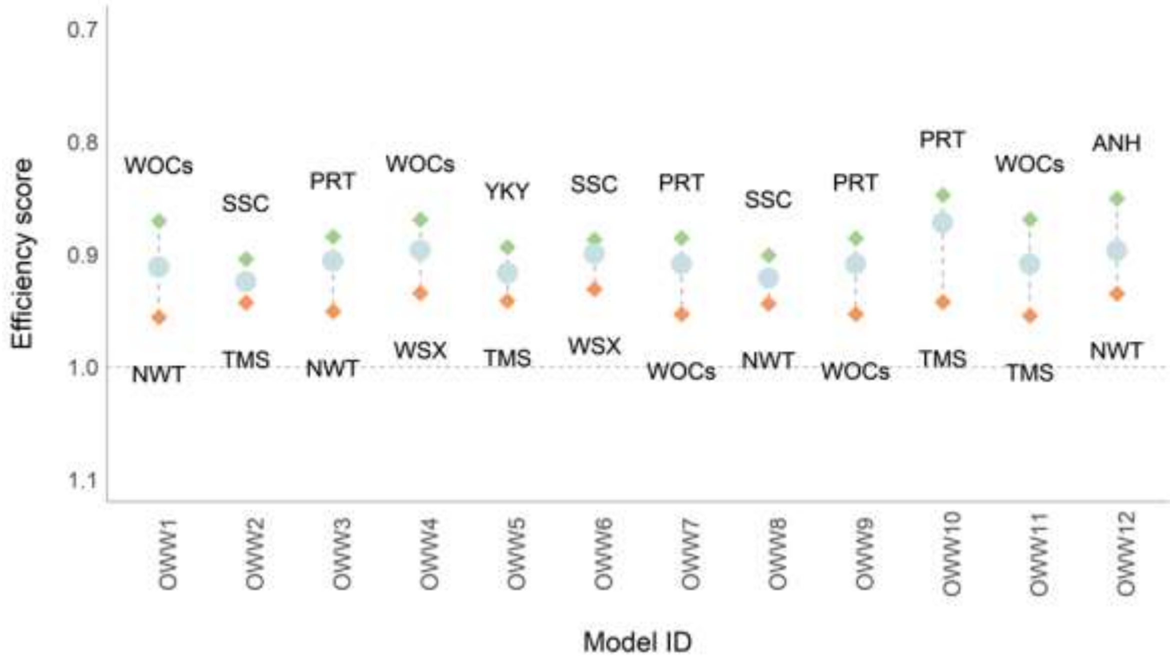


Figure 5. **Source type and economies of scale findings in water resources plus models are sensitive to the inclusion of Portsmouth and small WOCs more broadly**



The inclusion of WOCs, and in particular Portsmouth and small WOCs, has a large impact on key model outputs used in setting cost baselines and the efficiency challenge. Assessment of relative efficiency is affected by company inclusion, with an average range of 7 percentage points in wholesale water models. The upper quartile efficiency score varies by between 4 and 10 percentage points depending on the model in question, as shown in 0. A range of 4 percentage points around the efficiency challenge is worth over £620m in 2016/17 CPIH prices over the course of AMP7, with 10 percentage points worth £1,550m. Sensitivity is greatest around the inclusion of Portsmouth, WOCs, Thames and United Utilities. UQ efficiency score is sensitive to excluding individual companies, and in particular the WOCs in Ofwat's wholesale water models



Note: *Upper quartile challenge calculated based on PR14 methodology; labels show the identity of the company which leads to the highest and lowest efficiency challenge when excluded, for each model*

This evidence casts doubt on the validity of efficient cost estimates derived from Ofwat's water models, particularly for very small and very large companies. A lack of robustness to company inclusion or exclusion suggests that many models and variable coefficients are highly dependent on individual observations. This has sizeable impacts on key price control outputs such as company and efficiency scores across the industry (and hence predicted cost baselines).

Whether or not to include variables within a model should follow careful consideration of outlier analysis (such as we are presenting here), in order to better understand the robustness of model coefficients. Variables which are highly sensitive to the inclusion of individual companies should be treated with caution, especially where they behave as company dummies (discussed further in the next section) representing individual company operating environments.

Predicted cost baselines and efficiency challenges which explicitly account for noise and a lack of robustness are likely to improve customer and company outcomes at PR19. Simplistic use of model outputs, without considering insight from how reliable those predictions might be, can lead to arbitrary redistribution of allowances from large to small companies (or vice versa), or between companies with opposing characteristics (i.e. being "large" or "small" on any variable). This can lead to an excessively stringent efficiency challenge being applied, if it resulted in an arbitrarily large distribution of model residuals.

2.2 Models should avoid variables where there is little variation (cf. criteria 3)

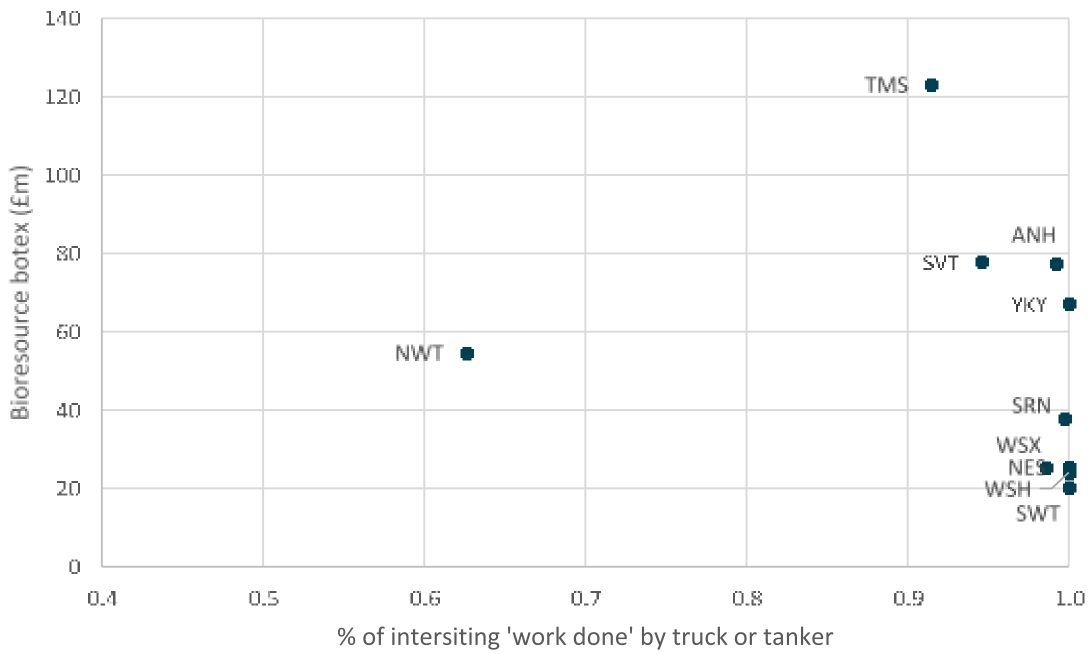
An extreme version of the issues outlined in the previous section arises when most of the industry variation for a particular variable occurs within one company. In that case an econometric model will treat that company (in effect) as a dummy variable. What then happens, the econometric model will simply act to pass residual value difference into that co-efficient – i.e. it does not predict a reliable cost relationship – it simply ascribes a relationship based on whether the company's costs are above or below average – i.e. it assumes all of that residual variation is due to that one factor. This will likely result in that factor having a coefficient with the wrong magnitude, and perhaps even the wrong sign, depending on whether the company's historic costs are above or below average.

An example of this is in bioresources. Both Ofwat and other companies have suggested the inclusion of variables for which UU has the vast majority of industry variation, for example:

- % of sludge disposed to farmland – UU is the only company with a material quantity of sludge disposed of by incineration (due to historic risk management of landbank availability)
- % of intersiting work done by truck/tanker – UU is the only company with a sludge pipeline. One company also suggested "Average distance intersited by pipeline", which is equivalent.

Figure 7 illustrates that the cost driver "percentage of intersiting work done by truck or tanker - this data is sourced from the industry datashare and is averaged over the six year period.

Figure 7 - Intersiting 'Work Done' by Truck or Tanker



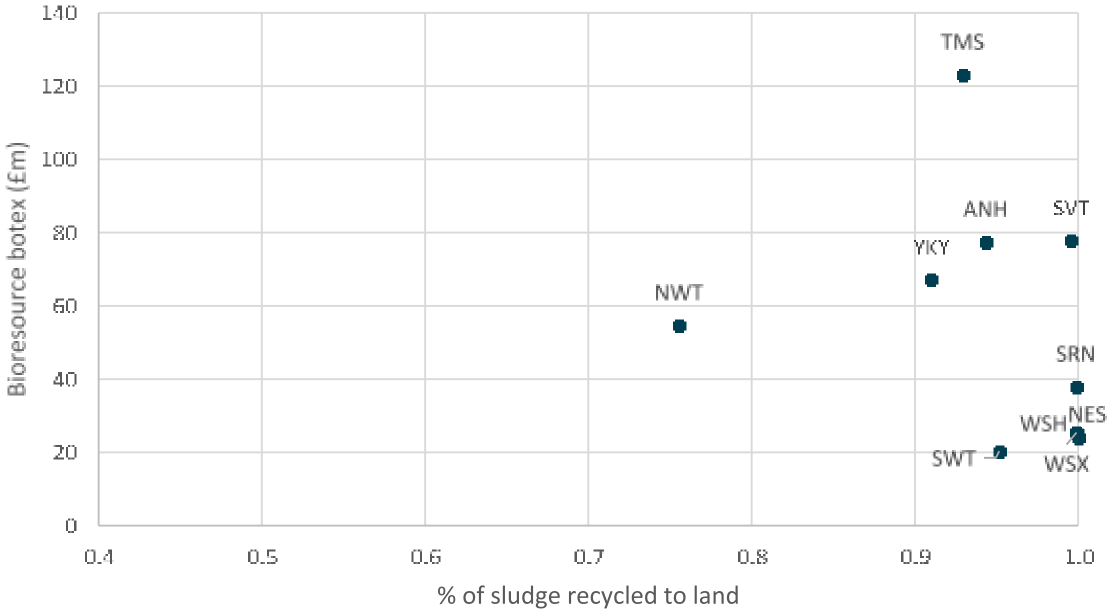
We can test the extent to which this variable is acting as a United Utilities dummy by dropping United Utilities from Ofwat's Bioresources models. The table below presents the results of this analysis.

	OBR1	OBR2	OBR3	OBR1 - no UU	OBR2 - no UU	OBR3 - no UU
ln(Connections)	1.002***			0.993***		
% of work done with truck or tanker	0.020***	0.017***	0.019***	**0.022**	**0.032***	**0.016**
% disposed to agricultural land	-0.021***	-0.018***	-0.018***	-0.031***	-0.024***	-0.027***
ln(Sludge produced)		0.940***	0.912***		0.981***	0.880***
ln(Total intersiting 'work done')			0.061			0.074
Constant	-3.728***	-0.554	-1.014*	-2.92	-1.621	0.149
N	60	60	60	54	54	54
Adjusted r squared	0.862	0.878	0.88	0.878	0.889	0.89
VIF statistic	2.02	1.99	1.98	2.581	2.798	5.963
RESET p value	0.012	0.003	0.002	0.012	0.008	0.003

Given that UU's costs relative to that variable appeared to be average, it is not surprising that the coefficient's magnitude remains similar once UU is excluded, however the cost driver (more correctly) loses its statistical significance. This suggests that the majority of the statistical fit is being provided by the one company acting as a dummy variable. We do not consider this to be a robust way to identify and evaluate differences in costs between companies.

Similarly, one of the cost drivers in Ofwat's Bioresources models - % disposal to farmland - could be considered a dummy variable for United Utilities. This can be seen in Figure 8 below. These factors should not be included in any econometric cost model for Bioresources (or in aggregated models which include Bioresources) because the model relations ship will likely be unduly influenced by that one observation.

Figure 8 - Disposal to farmland across the industry



2.3 Models should contain factors which represent the broad range of service delivery (cf. criteria 1)

There are many reasons for differences in efficient costs between companies. These arise due to differences in the environment, or the long term asset inheritance of companies.

For the Wastewater service, the work prepared for us by Arup and Vivid reviewed this in detail[1]. This considered an initial long list of factors and narrowed these down to the few most important areas to have an appropriate cost driver for within a cost assessment model.

For the water service, there are many different sources of such variations, and it is important that the specification of cost assessment models should reflect all areas, particularly when different areas have different cost characteristics.

For example, many of the water models proposed rely on one or more pumping related variables. Whilst pumping can be a material activity and accounts for a large proportion of power costs, it is less obvious that it is key driver of <u>differences</u> in cost between companies, any more than other factors, beyond variations due to scale. Also, although pumping costs may also be material to operating costs, they are less material to overall totex. It is also the case that gravity fed water resource systems have higher capital costs due to maintenance of reservoirs and aqueduct systems. Further downstream into water treatment, surface water sources typically also require higher levels of treatment than groundwater, and hence more capital maintenance and more "other opex".

Mitigants of this for construction of a reasonable econometric model could be:

- To ensure that models reflect a sufficient range of factors which capture all material sources of variation between companies, in order to more fairly reflect the costs of all companies, rather than

[1] Arup and Vivid Economics (2017) *Understanding the exogenous drivers of wholesale wastewater costs in England and Wales*

only including one and hence biasing cost assessment toward one group of companies at the expense of another;

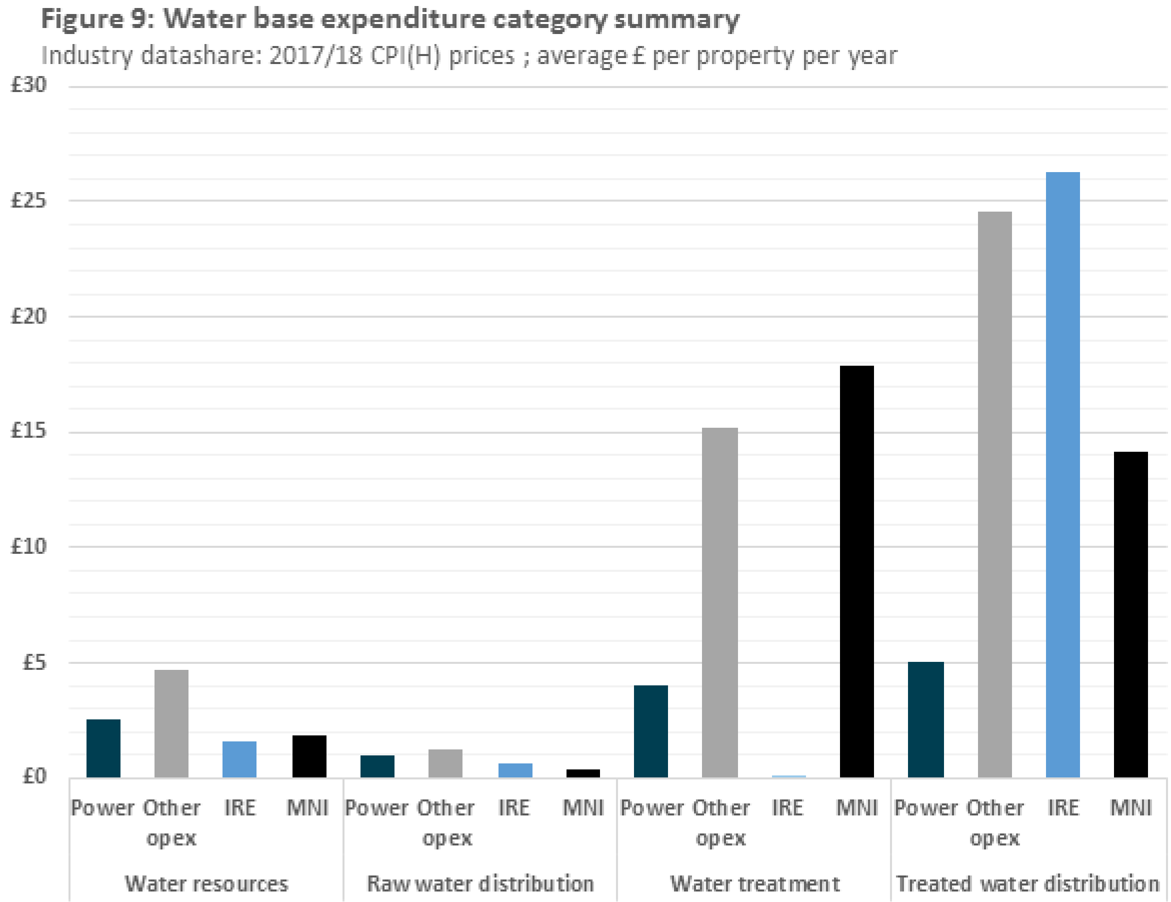
- To identify potentially opposing cost drivers, and include both or neither – as inclusion of only one could (again) unduly bias results; or
- Analyse and visually inspect the data to ascertain if relationship with costs is reasonable, and/or if some companies appear to have outlying characteristics, then those might better be addressed via cost adjustment claims than within an econometric model.

In the following two sub-sections, we present differences in component unit costs between companies across the water and wastewater value chains. This enables observation of which areas of cost result in the largest (non-scale) variation in cost, and hence provides some insight into the key areas which econometric models should be seeking to explain.

2.3.1 Water value chain analysis

When developing a suite of suitable benchmarking models, it is important to understand the activities and resulting expenditure that underpin each of the value chains. Selected explanatory factors should not simply aim to explain major cost items, as that should be fulfilled by the scale variable, but to explain significant variations in unit costs between companies. This analysis is helpful as it can highlight if undue weight is being placed on certain factors, causing incorrect reallocation of expenditure between companies when a basic cost threshold is generated.

Whilst the absolute cost alone does not necessitate the inclusion of a factor within a model (since if there is no variation then this will be explained by the scale driver), normalised data can quite quickly highlight where the primary areas of spend occur across the industry. This can then form the basis for investigating which areas of the value chain - and therefore which explanatory factors - should most appropriately be used within a benchmarking model.



Error! Reference source not found. above provides a useful starting point as it helps to identify which areas of each value chain are most significant[2]. Whilst within each of the individual value chains there can be different specific factors, in general, power typically represents pumping activities or advanced treatment technologies, IRE (infrastructure renewals expenditure) represents maintenance activities associated with underground assets, aqueducts and impounding reservoirs, and MNI (non-infrastructure maintenance) represents the maintenance of civils structures and mechanical & electrical equipment (treatment works, pumps etc.)

Other opex will hold all other direct and indirect operating expenditure. Whilst other opex largely reflects expenditure associated with labour, chemicals and repair activities that are not capitalised, it can also reflect more variable company activities (e.g. the cost of engineering inspections of impounding reservoirs).

[2] Using definition of botex consistent with that within section 2 of Ofwat do-file "ofwat_water_models".

Figure 10: Water Resources base expenditure category summary
Industry datashare: 2017/18 CPI(H) prices, average £ per property per year

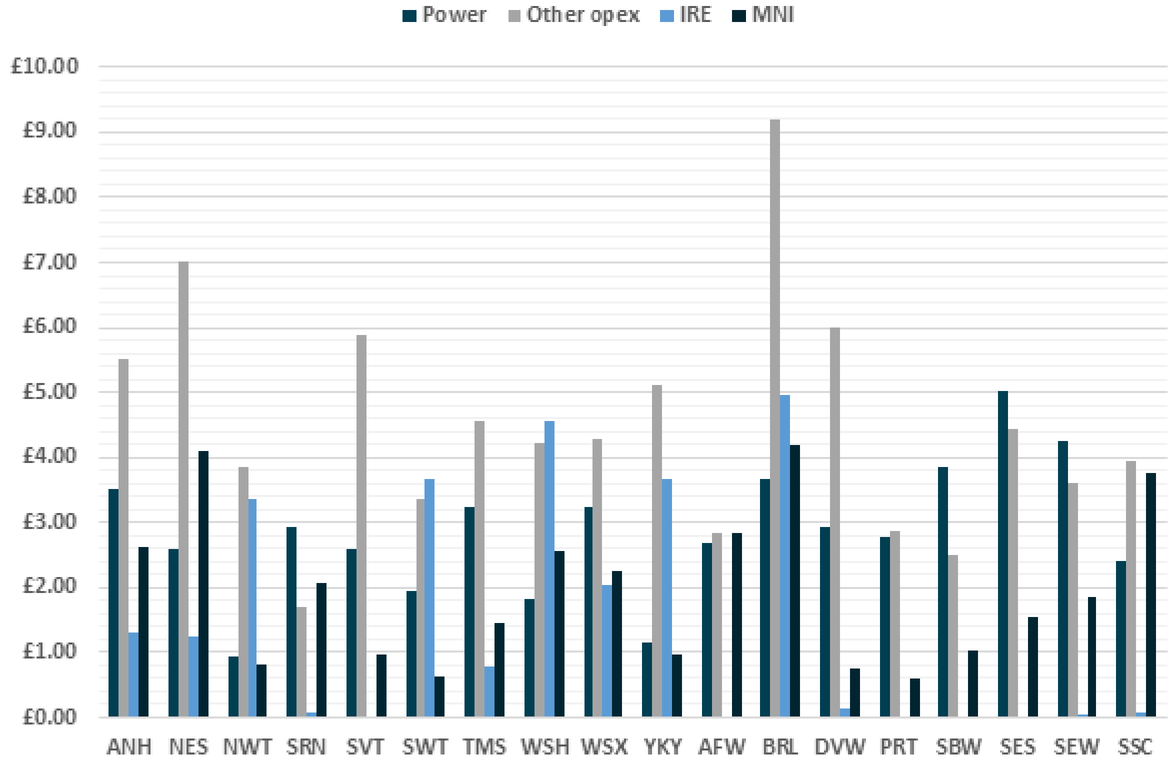


Figure 11: Raw water distribtion base expenditure category summary
Industry datashare: 2017/18 CPI(H) prices, average £ per property per year

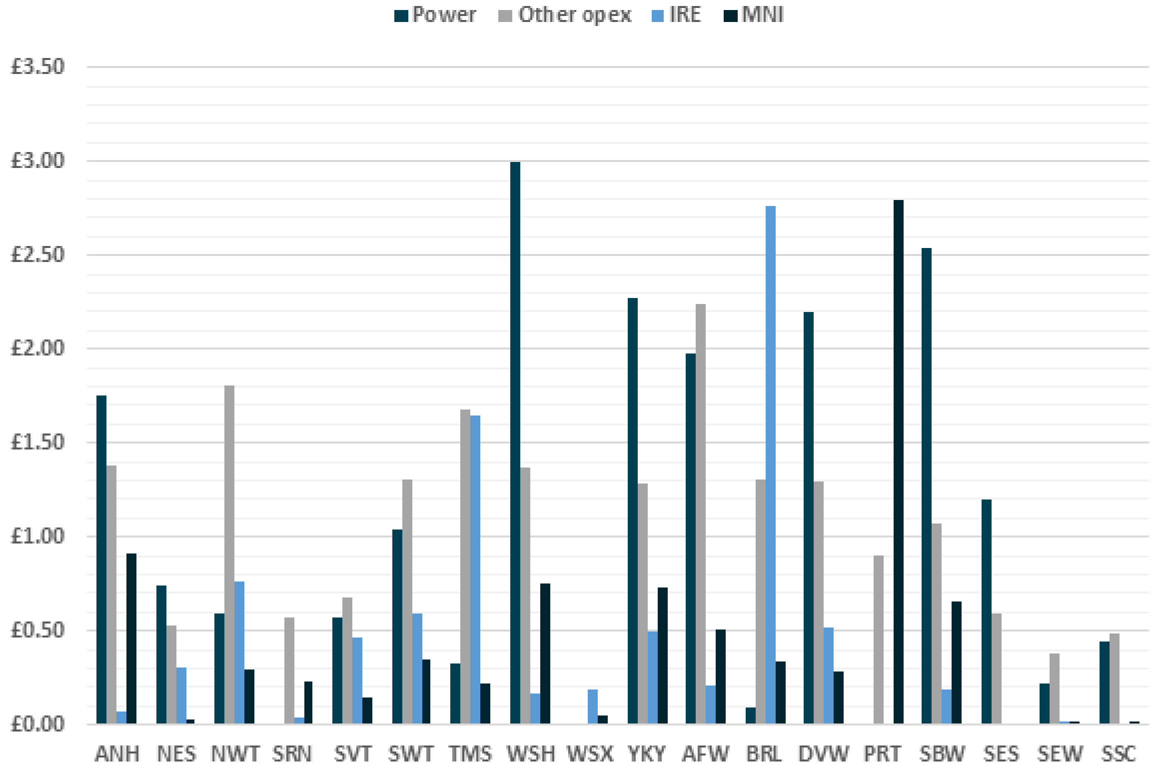


Figure 12: Water Treatment base expenditure category summary

Industry datashare: 2017/18 CPI(H) prices, average £ per property per year

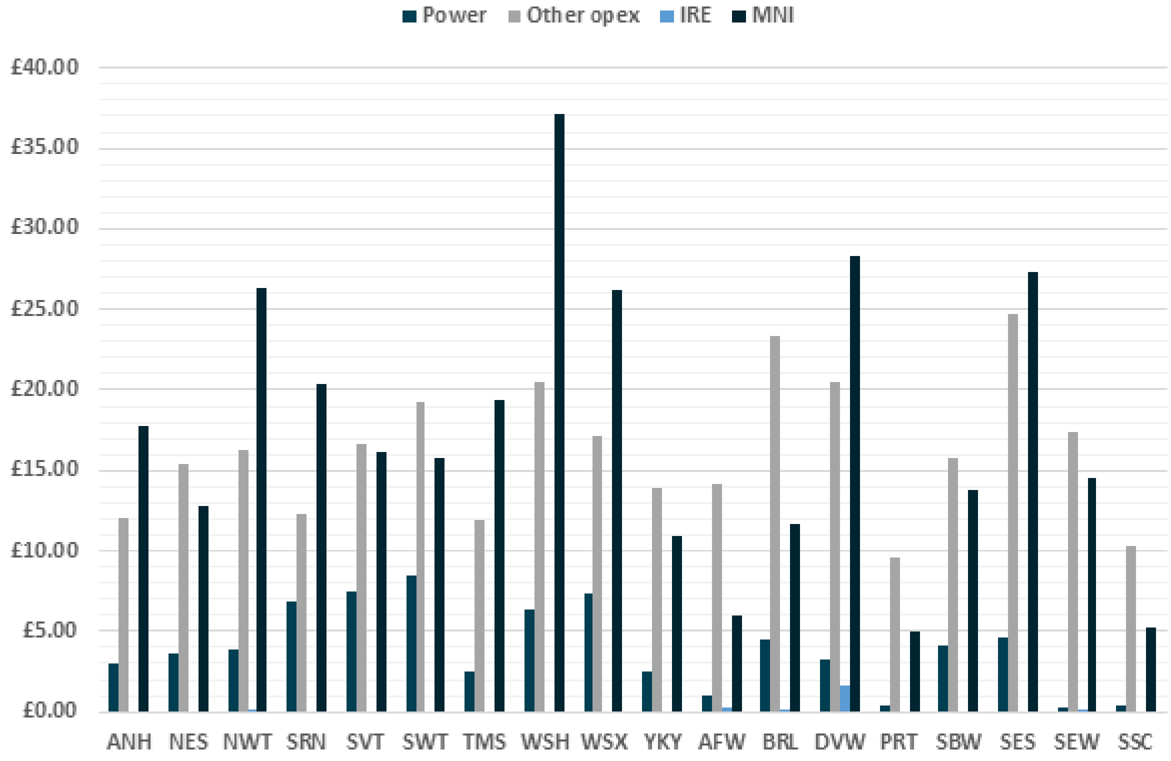
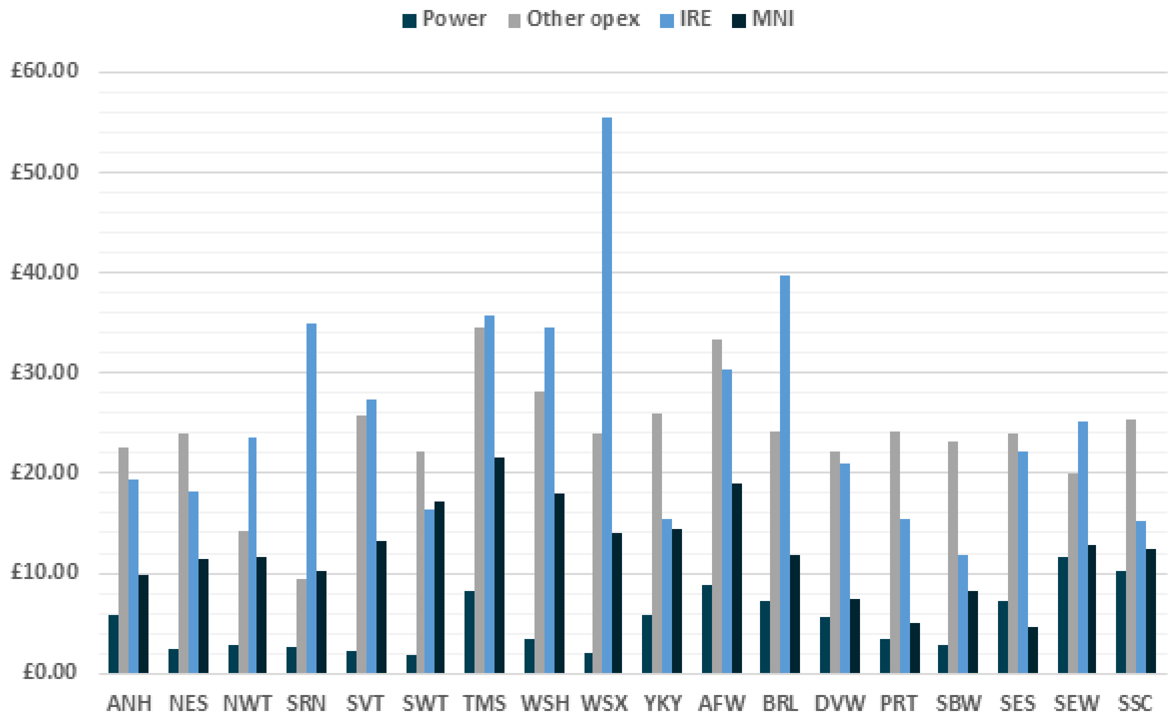


Figure 13: Treated water distribtion base expenditure category summary

Industry datashare: 2017/18 CPI(H) prices, average £ per property per year



2.3.2 Wastewater value chain summary

Findings from Arup / Vivid analysis

The research undertaken by Arup and Vivid[3] focused on fourteen narratives (having considered more than 200 narratives overall) where the potential to add value was greatest. An initial filtering of narratives tested the potential for improved understanding of the relevant areas to improve cost assessment. This considered, taking account of all company operating environments in England and Wales, whether the narratives described processes that had significant effects on costs, whether they described processes that varied between companies, and the degree of existing comprehension of the process and data available on relevant company conditions. On this basis, the project concentrated on fourteen narratives, encompassing all aspects of companies' activities - the ones in bold text were taken forward as the most valid and comprehensive ones to progress into models of wastewater costs, ensuring that models reflected variation in both quality and quantity. The regional wage actor was later rejected for inclusion in an econometric cost model, in part due to it only being an issue for a single company, and hence was deemed more suitable as a candidate cost adjustment claim.

[3] Arup and Vivid Economics (2017) *Understanding the exogenous drivers of wholesale wastewater costs in England and Wales*

Factor	Driver	Maturity of Narrative	Data Quality at National Level
Annual amount, intensity or frequency of rainfall (runoff)	**Quantity**	**High**	**Fair to good**
Influence of topography on peak flows in networks	Quantity	High	Fair to good
Load received or removed	**Quality**	**High**	**Variable – poor to good**
Industrial loads received or removed	Quality	High	Poor to fair
Environmental designations and planning conditions	Other	High	Fair
Regional wages	**Economic**	**High**	**Fair**
Sparsity and urbanisation	**Density**	**Medium**	**Fair**
Sludge land bank	**Sludge**	**Medium**	**Fair**
Flood risk	Quantity	Medium	Fair
Hidden or culverted rivers contributing to sewer flows	Quantity	Low to medium	Poor to fair
Asset accessibility	Density	Low	Poor
Customer characteristics	Quality	Low	Poor
Groundwater infiltration or minewaters contributing to sewer flows	Quantity	Low to medium	Poor
Asset age	Quantity	Medium	Fair

Further assessment assembled a research strategy for each narrative. This revealed that some of these factors - such as the effect of 'hidden' subterranean urban rivers on flow volumes - described processes that were not sufficiently measurable or did not materially impact costs and so were not pursued in detail. In other areas - such as the relationship between load and costs - the nature of the relationship was already understood, but valuable insights could be gained through assessing the reliability of the data used in the sector.

The data collection and modelling work focused on the drivers identified as measurable and material. In the areas of drainage, treatment quality, economies of scale, urbanisation, and rural sparsity, the project assembled: detailed engineering evidence of how aspects of a company's environment affect its assets, operations and costs; summary data on how these factors vary between company regions; and modelling evidence on the extent to which factors affect costs. This produced positive recommendations for modelling in the areas of drainage, economies of scale and urbanisation. For treatment quality, the engineering case for materiality is also set out but more complete and consistent sector-wide data on discharge permits is needed to account for inter-company variation in benchmarking models. For sparsity, though the engineering narrative developed suggested costs should be higher in very sparse areas, this could not be substantiated by modelling evidence.

Economies of scale are most pronounced at the level of treatment assets (not at the overall company level). This is because economies of scale occur when the unit cost of service provision falls as the volume of service increases. Previous approaches, including that used at PR14, focused on economies of scale at the level of companies, which would imply the existence of managerial or operational efficiencies that can be exploited by larger companies but not by smaller ones. However, for sewerage companies these efficiencies are modest in comparison to economies of scale at the level of assets, where the average cost of wastewater treatment declines as the size of a treatment works increases.

Arup's engineering assessment indicated that unit costs of treatment in works of 1,000 population equivalent (PE) capacity are as much as 6 times greater than costs at larger (25,000 PE) works, while industry data shows disparities between the volume of treatment carried out at small works by different companies (1 – 7 per cent). In econometric models, the percentage of load treated in small works shows a consistently positive, significant relationship with costs. Hence, the study recommended using this variable to account for economies of scale in future models.

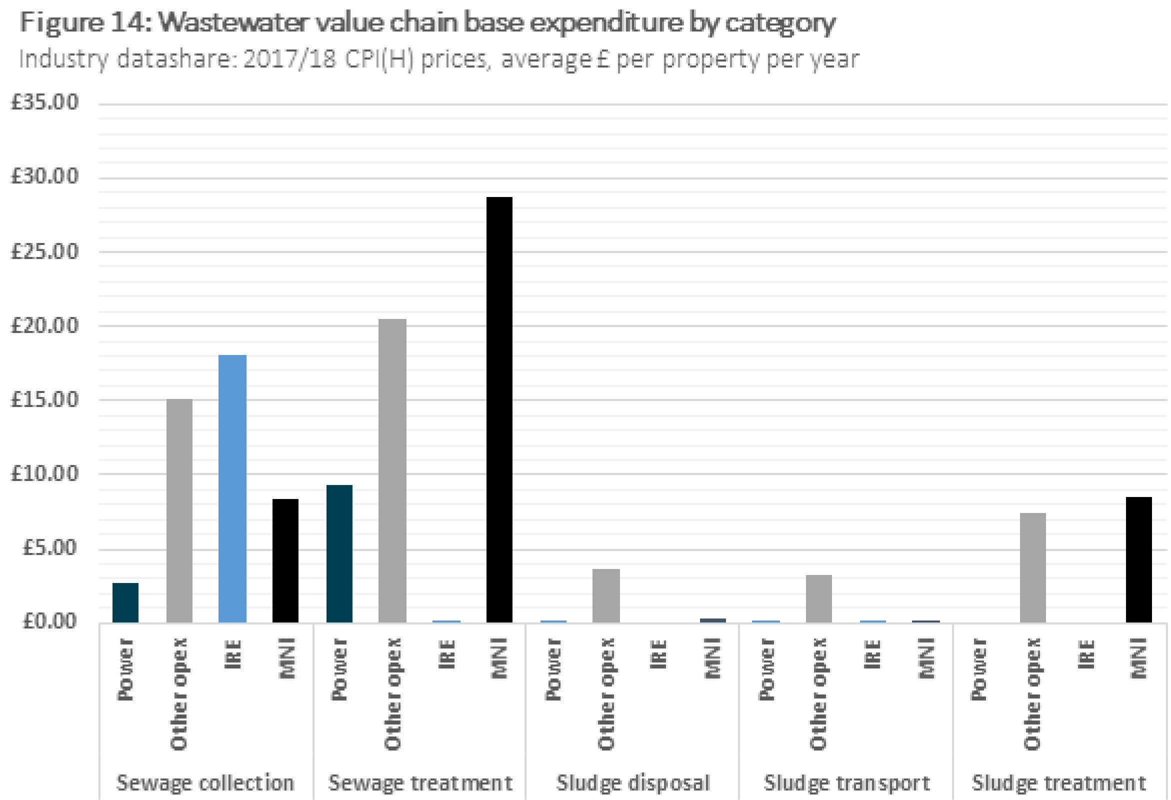
Drainage costs be accounted for using data on urban runoffs. Drainage is a significant component of the wholesale wastewater service, but the relationship between company-level data and costs had not previously been explored. A company's activities in drainage service provision depend chiefly on inflows into combined drainage and sewerage networks. In general, the greater the volume of such inflows, the larger network and storage assets need to be, and the greater the amount of pumping. Data on urban runoffs, a proxy for drainage inflows, shows variation between company regions of the order of 50 per cent. Engineering analysis of the drainage processes and industry cost data suggest that costs associated with the drainage service vary notably between companies. This hypothesis is supported by modelling evidence showing that urban runoff does have a positive, significant relationship with costs. On this basis, the project recommended including urban runoff as a driver in cost benchmarking models.

The provision of services in urban areas can be costlier than elsewhere for a variety of reasons. Access to networks for maintenance is limited by hard surfaces and the need for permissions for lane closures, operations are hampered by slow traffic speeds, while treatment assets may be constricted by land footprints and more stringent conditions on odour. However, quantitative evidence on the relationship between urbanisation and costs remains limited: for example, the network density variable used in the PR14 models measures the number of connections per kilometre of network averaged across both urban and rural parts of a company's region, a poor proxy for the relevant narratives. A more accurate reflection of how much of a company's activities take place in urban environments is the proportion of treatment capacity located in districts classified as urban by the ONS. Though this variable remains an imperfect proxy, it has a positive, significant relationship with costs in econometric models. The study therefore recommended using this variable, or one similar to it, in future models.

Treatment permits affect treatment unit costs, but current sector-wide data is insufficiently complete to value this in benchmarking models. The level of discharge permits drive the choice of treatment technologies, which in turn can have an appreciable effect on unit costs of treatment. For a large works, moving from a 'basic' permit of 20 mg/l BOD5 to a more 'stringent' permit that limits discharges to 3 mg/l NH_3 requires the installation of a nitrifying activated sludge process, a shift that causes unit costs of treatment to increase by around 47 per cent. The stringency of permits varies markedly between treatment works and would thus be expected to be an important factor in explaining differences in costs. However, complete data on the permits at different treatment works is not currently reported across the industry. The study recommended this be addressed through more extensive sharing of information across the sector.

The results of this work remain valid and hence we maintain support for our proposed suite of wastewater models, which have been the result of a material exercise in reviewing the drivers of cost, based on extensive review of both engineering narratives and economic analysis.

2.3.3 Wastewater value chain analysis



Whilst within each of the individual value chains there can be different specific factors, in general, power typically represents pumping activities or advanced treatment technologies, IRE (infrastructure renewals expenditure) represents maintenance activities associated with underground sewer assets, and MNI (non-infrastructure maintenance) represents the maintenance of civils structures and mechanical & electrical equipment (treatment works etc.)

Figure 15: Sewage collection

Industry datashare: 2017/18 CPI(H) prices, average £ per property per year

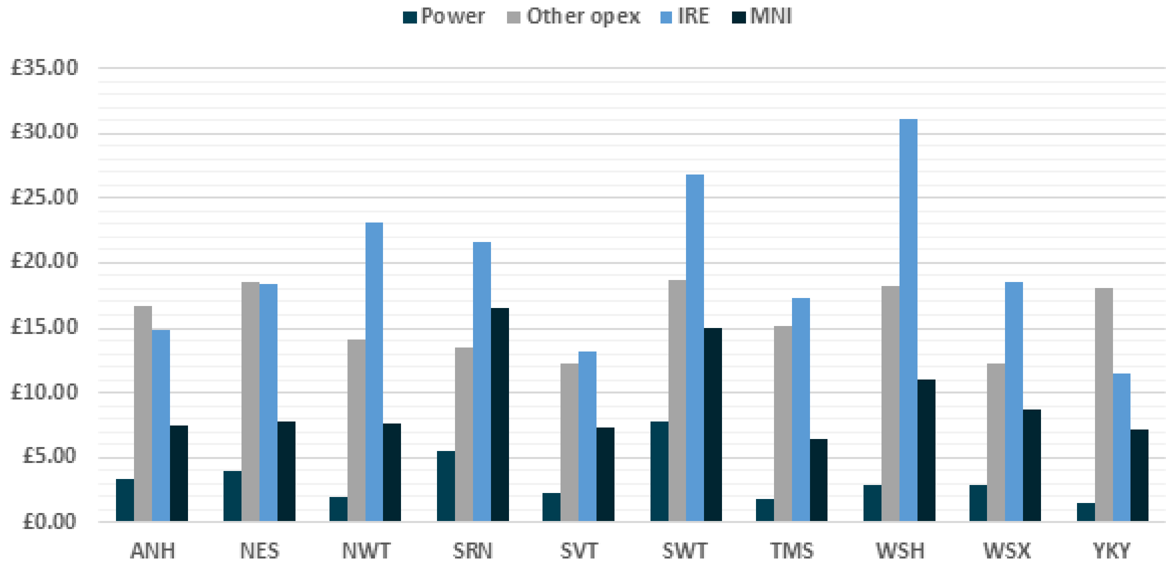


Figure 16: Sewage treatment

Industry datashare: 2017/18 CPI(H) prices, average £ per property per year

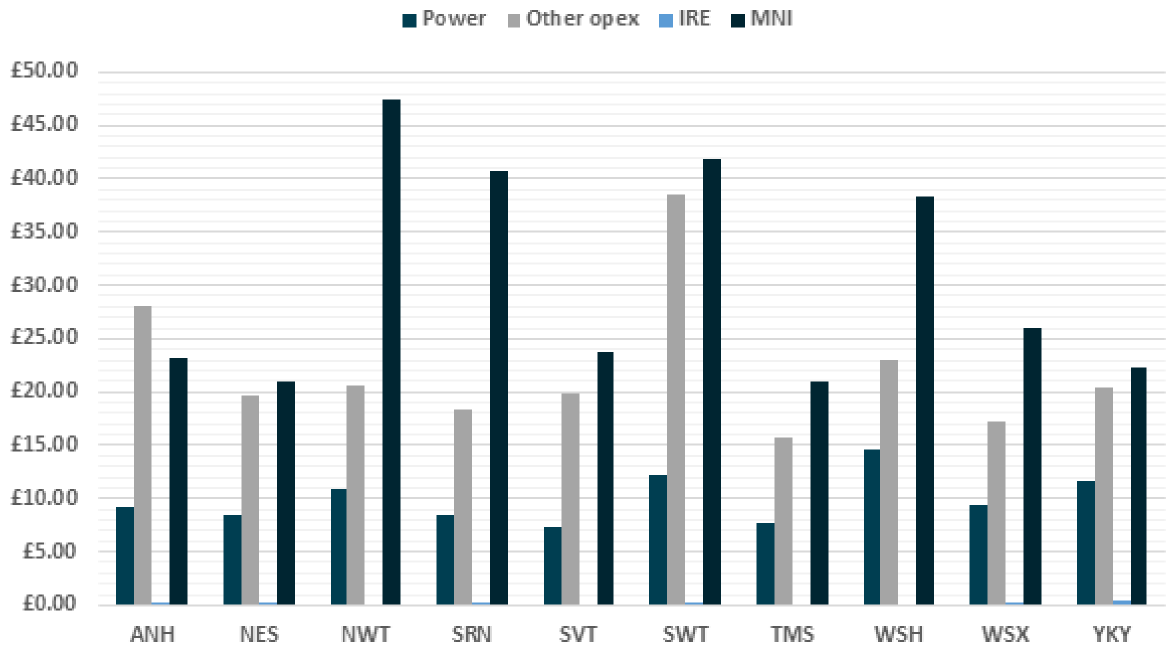


Figure 17: Sludge transport
Industry datashare: 2017/18 CPI(H) prices, £ per tds

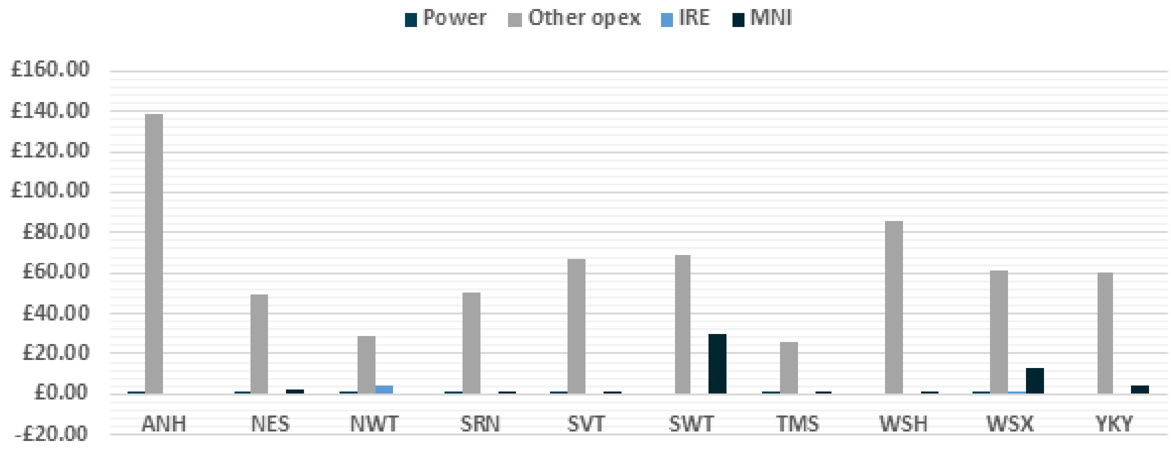


Figure 18: Sludge treatment
Industry datashare: 2017/18 CPI(H) prices, £ per tds

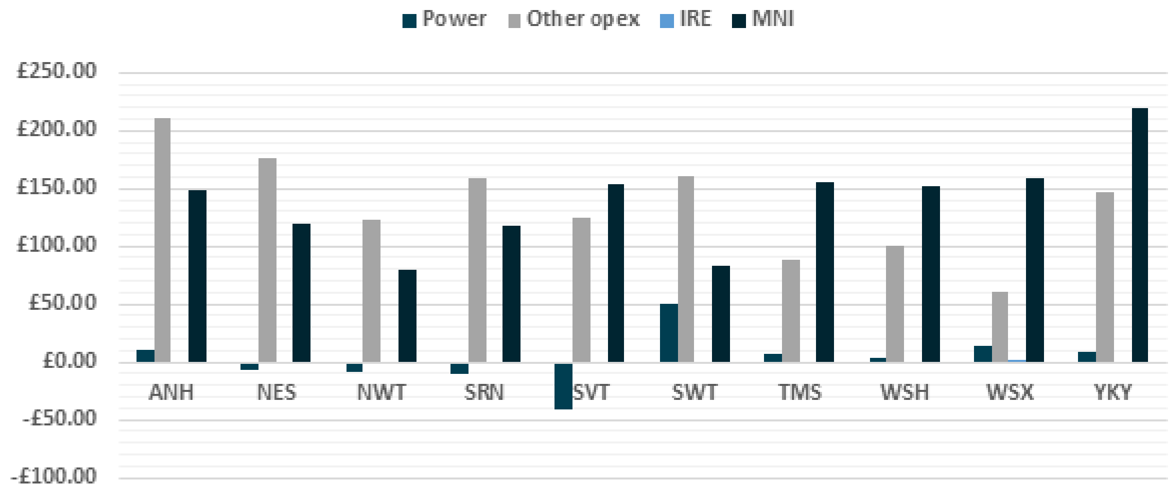
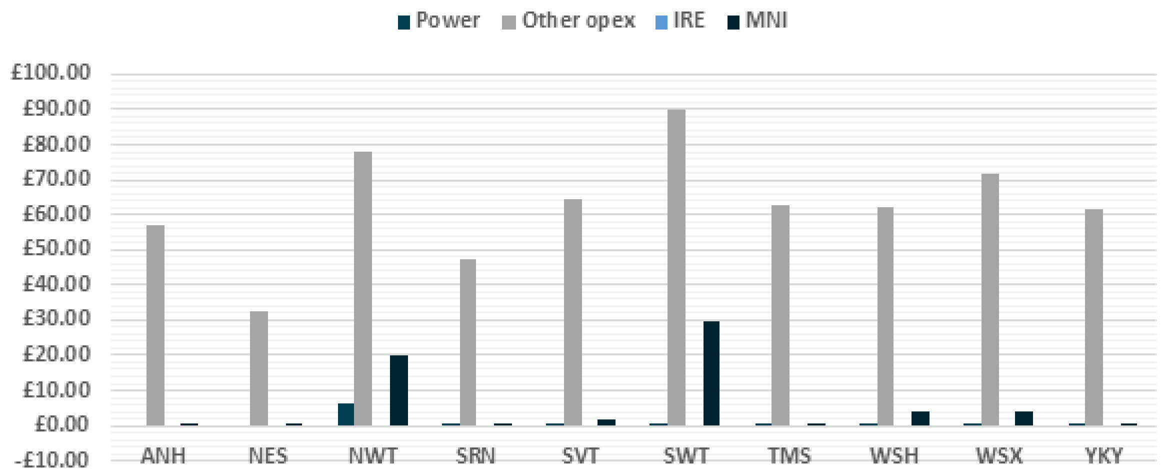


Figure 19: Sludge disposal
Industry datashare: 2017/18 CPI(H) prices, £ per tds



2.4 Other considerations for constructing cost assessment models

2.4.1 Models should avoid explanatory variables that correlate with scale, even where this is only relatively weak (cf. criteria 1 and 2)

Given the very small dataset, it is highly likely that some explanatory variables will reflect scale in a non-causal way – for example, if WoCs tend to have different operating characteristics to many of the larger regional WaSCs, such as a greater reliance of borehole sources than surface water / reservoir sources. This risks coefficients of variable being overvalued if there is any (even minor) collinearity between variables used and scale. This might unduly skew cost estimates for some companies to the detriment of others. One way to test this is to observe the impact on the scale variable co-efficient following inclusion / exclusion of a particular factor. If the scale variable co-efficient falls, then it may useful to test for any correlation between scale and the added variable.

2.4.2 Prior assumptions need to be tested with companies (cf. criteria 2)

It would be prudent to test any *a priori* assumptions with companies, before they are applied within a particular cost assessment. For example in the Water Resources models, Ofwat states "*A number of companies present models with a positive coefficient on the proportion of water from reservoirs. We question whether this is the expected sign in a water resources model.*" We do not view this as particularly surprising since we do not believe it should be unexpected that "proportion of water from reservoirs" would have a positive co-efficient (and conversely that "proportion from boreholes" would be negative). Whilst there are some costs (largely related to pumping) which arise for borehole sources, which may be avoided for companies with a high proportion of reservoir sources, there are also additional costs resulting from impounding reservoirs (such as maintaining the safety of impounding reservoir structures through detailed engineering inspections and further remedial maintenance actions) which are avoided through borehole sources. These additional costs, in our experience, can outstrip the cost of additional pumping associated with borehole sources.

We have submitted a draft cost adjustment claim on 3 May, which sets out more evidence on this point in more detail. In general, we suggest Ofwat shares prior assumptions - or seeks specific evidence to validate them - before applying them to any specific cost assessment.

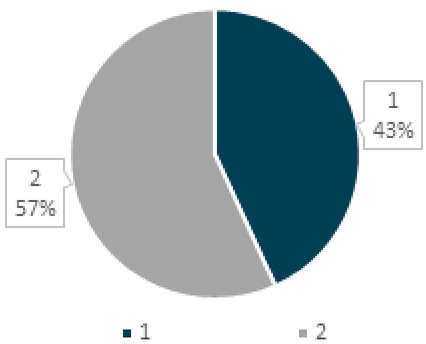
3 Comments on Wastewater Models

3.1 Wastewater Treatment Quality

We note that Ofwat only includes a measure of wastewater treatment quality across two wastewater network plus models. A number of other companies propose alternative measures of treatment quality across wastewater models. We consider the inclusion of both surface water drainage and wastewater treatment quality to both be essential and integral to any robust wastewater cost assessment model.

The chart below shows the composition of wastewater network plus costs. It's clear that sewage treatment contributes significantly to overall network plus costs. Omitting a factor from such models that is widely accepted as a cost driver in the wastewater value chain could understate treatment costs and therefore underestimate the efficient costs for companies with relatively stringent consents. Therefore, we consider Ofwat should include a measure of treatment quality across all wastewater models (with the exception of sewage collection and bioresources models).

Figure 20: Composition of Wastewater Network Plus Costs



source: industry datashare (16/17 prices)

Our analysis indicates that the addition of treatment quality measures to network plus models also improves performance. We tested adding two alternative measures of treatment quality to Ofwat's network plus models ONP1-ONP8: percentage of load with Ammonia consent less than 1mg/l, and percentage of load with Ammonia consent less than 1mg/l and BOD less than 10mg/l. Ramsey's RESET test was applied to each model run to test for misspecification. We picked these models as Ofwat did not include a quality driver.

The following table presents the RESET test p-values, averaged across the eight network plus models.

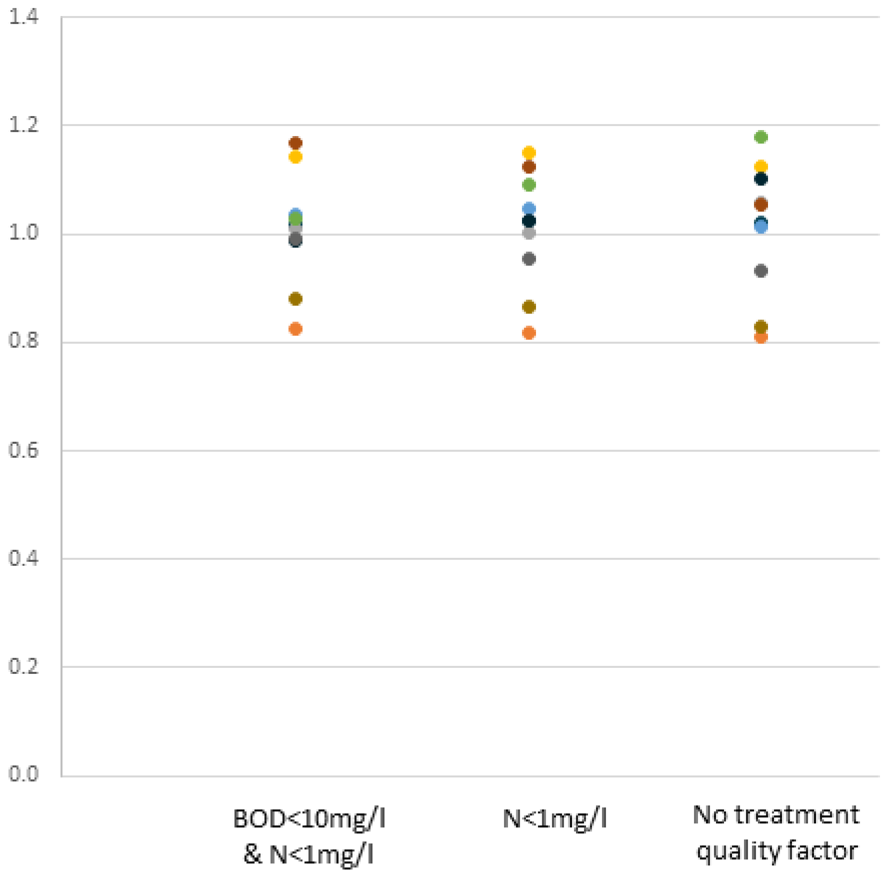
Quality Factor	Average
No quality variable	0.00725
% of load with N < 1 mg/l	0.0095
% of load with N <1 mg/l and BOD <10 mg/l	0.24425

Note: this table presents average p-values from Ramsey's RESET test. The average was taken across Ofwat models ONP1-ONP8. A value greater than 0.05 is generally taken to imply a 'pass'.

It's clear that percentage of load with Ammonia consent less than 1mg/l and BOD less than 10mg/l performs well compared to models that omit a treatment quality factor. There was also a marginal improvement in the adjusted r squared relative to the 'no quality variable' case when this factor was added. Additionally, the variable was highly statistically and economically significant across all models. We consider this to provide strong statistical evidence - in addition to a strong engineering rationale - that a measure of treatment quality should be included across all wastewater models. We note that a number of other company models have proposed this measure, which demonstrates there is considerable recognition across the industry of the case for its inclusion.

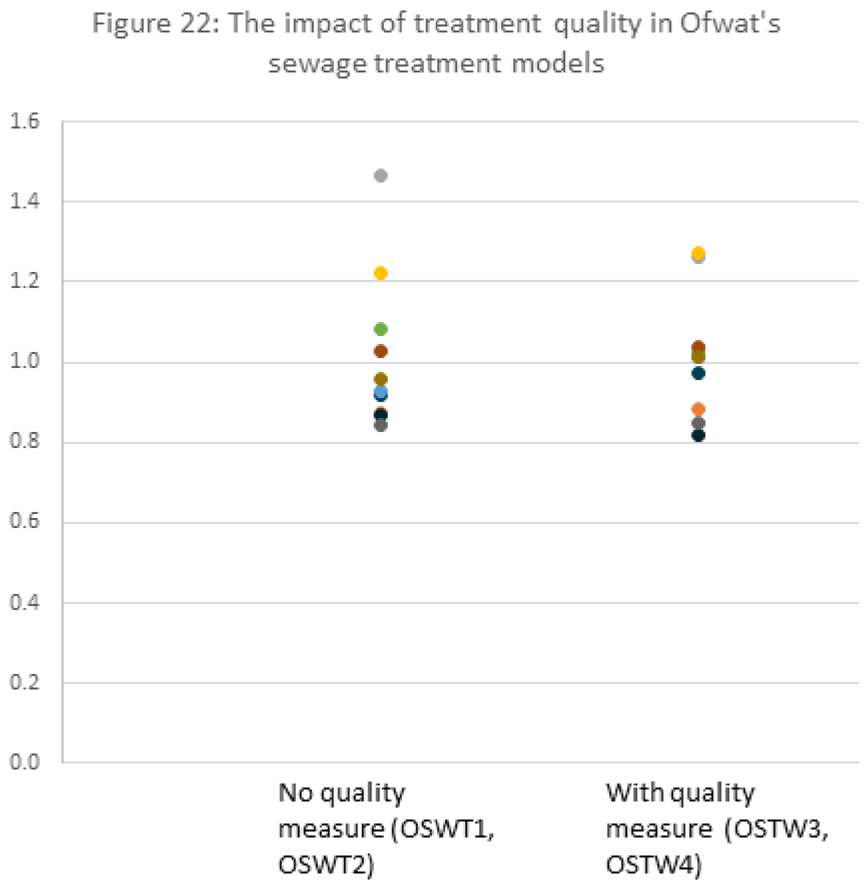
The diagram below demonstrates how the addition of treatment quality factors improves the residual spread. A smaller residual spread indicates the models better predict costs experienced in reality. The spread is calculated as the maximum residual minus the minimum residual.

Figure 21: Residual Spread Across Different Measures of Treatment Quality



Analysis of the underlying data indicates that adding percentage of load with Ammonia consent less than 1mg/l and BOD less than 10mg/l reduces the spread of residuals by seven percent, while adding percentage of load with Ammonia consent less than 1mg/l reduces the spread of residuals by ten percent. This further supports that that model fit is improved through the inclusion of a relevant treatment factor.

We can also see this effect in Ofwat's models of sewage treatment.



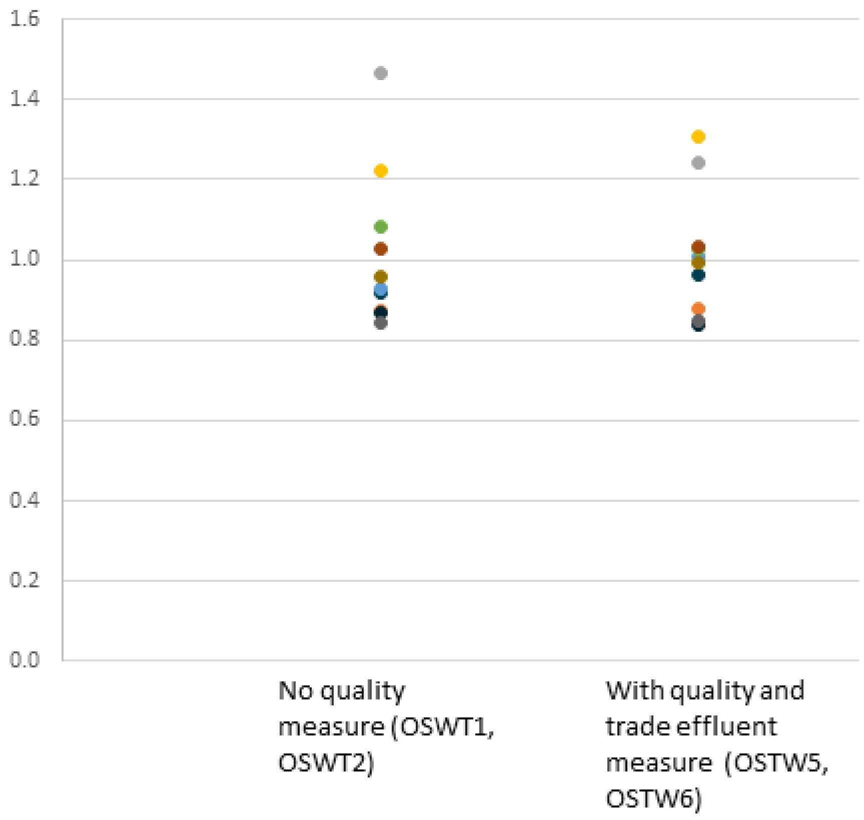
3.2 Trade Effluent

Trade effluent should be an important driver of wastewater treatment costs, given differences in industry between the different regions of England and Wales. Therefore, we support Ofwat's inclusion of trade effluent as a cost driver in its wastewater treatment models.

Given the strong engineering justification for its inclusion and the degree of variation across the industry, it is surprising that the factor does not have a stronger impact within the models. However, the coefficient is of the expected sign and the model fit improves once it is included, so we consider its inclusion is justified. The low impact of this factor is likely to be due to omitted variables bias – some other companies have relatively high unexplained costs within models, but low levels of trade effluent (for example Thames). The inclusion of such companies appears to be (in part) negating the expected relationship.

The scatter diagram below shows the difference in the residual spread between Ofwat's models which include and do not include trade effluent (some of the difference can also be attributed to the inclusion of treatment quality measures).

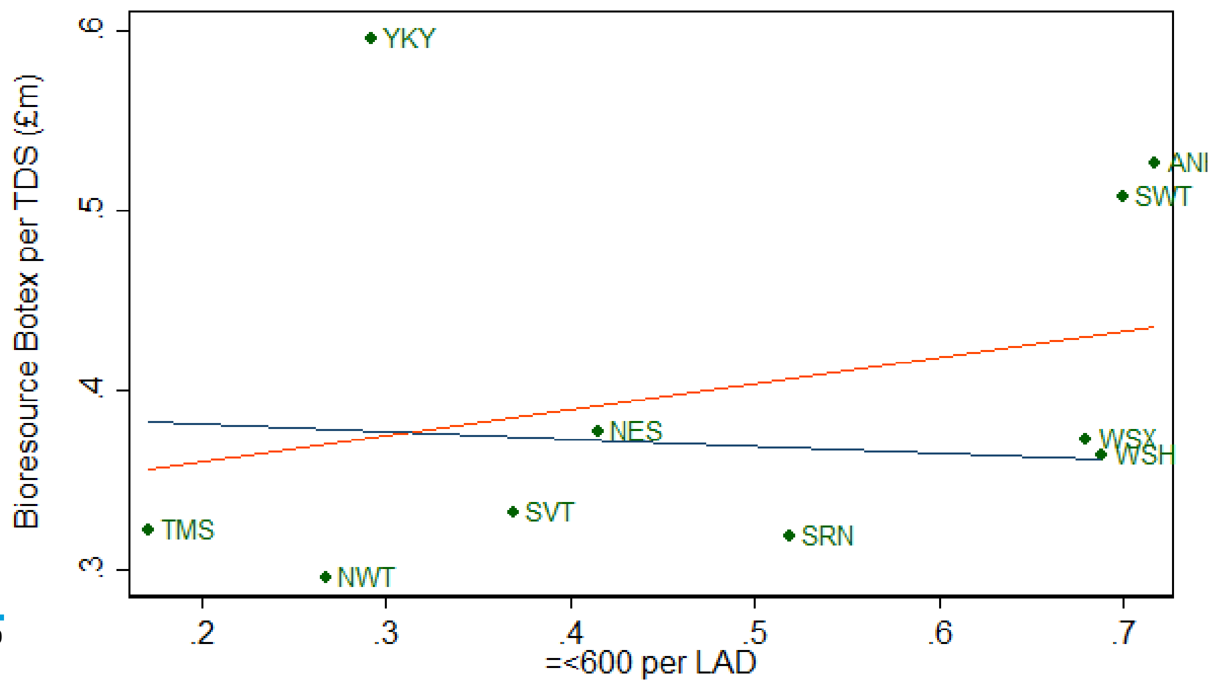
Figure 23: The impact of trade effluent in Ofwat's Sewage Treatment Models



3.3 Density and sparsity outliers

We note that a number of stakeholders have suggested that density measures should be included in Bioresources models. We agree that the relative concentration of population will drive variation in costs. However, we consider that Ofwat should be cautious in choosing a variable to reflect density. This is because there is the potential for some companies to exert high leverage within the model. This is shown in the following graph.

Figure 24 - Bioresources Unit Cost and Sparsity



The red line is the relationship between Bioresources unit cost and an Ofwat measure of sparsity. The blue line is the same relationship after dropping Anglian Water and South West Water. The fact that the relationship changes shows that these two companies have high leverage in models which include this measure of density. It is also not evident why the costs of Anglian and South West should be so much different from other similarly sparse companies (Welsh and Wessex), or why Yorkshire's costs are so much different from those of other companies. This all results in a lack of evidential support for the purported relationship between sparsity and cost.

4 Comments on Water models

4.1 Avoid exclusion of service components

4.1.1 Water models should include both pumping and source type

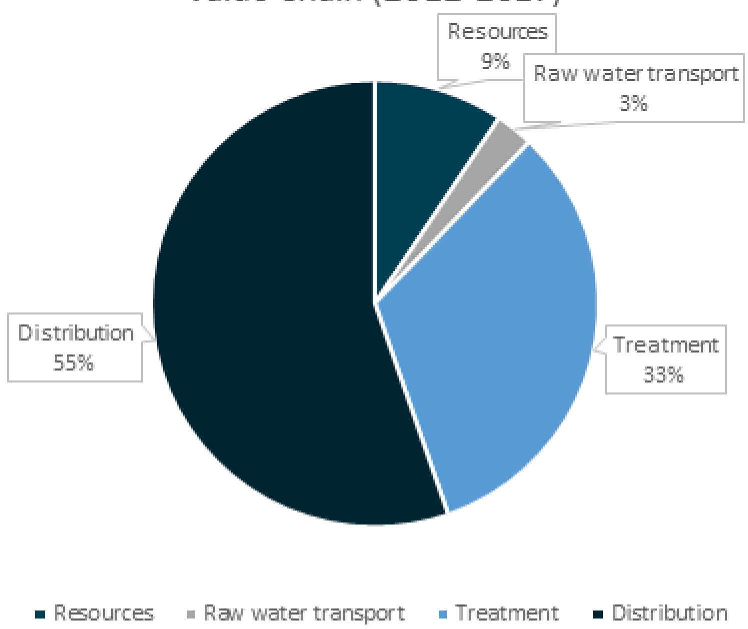
We consider that it is important for Ofwat to consider both treatment quality and pumping head in its water models. Given that surface water sources generally require higher treatment costs and ground water with increased pumping, it would not be correct to exclusively consider only one of these factors as a cost driver. Instead, Ofwat should recognise that each is associated with unique, and potentially offsetting, variations in cost and therefore it would be better for both to be included in any relevant cost model. Indeed, compared to the alternative of including just one of the two, it would actually be preferable to include neither - as this would produce a less biased result.

We would expect chemical costs to be higher per unit for surface water treatment works and for power costs per unit to be higher for groundwater treatment works. Both factors are also statistically significant when utilised in botex models. Excluding only one factor would unfairly detriment companies with a higher than average level of that factor.

4.1.2 Exclusion of raw water distribution in Wessex' model

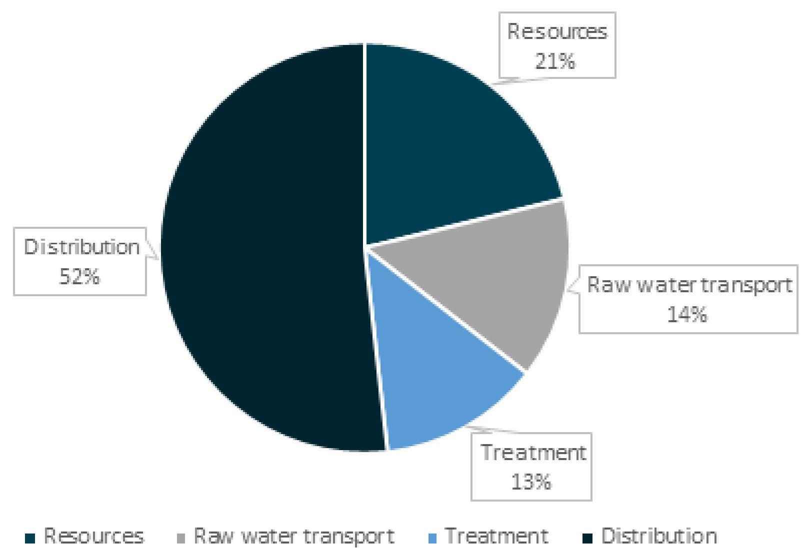
We note that the Wessex water resources model excludes raw water distribution costs. As a general point, we consider that excluding value chain elements introduces unpredictable distortions into econometric models. For example, Figure 25 demonstrates that three percent of water botex relates to raw water distribution, which is a fairly small percentage.

Figure 25: Share of Total Industry Botex Across the Water Value Chain (2012-2017)



However, Figure 26 demonstrates that the industry is faced with more significant pumping requirements in raw water distribution than in water treatment. It's not clear what aggregation of pumping head has been included, however it's clear that a robust model relies on the underlying data being accurate. Omitting relevant costs would seem to undermine this.

Figure 26: Composition of Total Industry Pumped Head

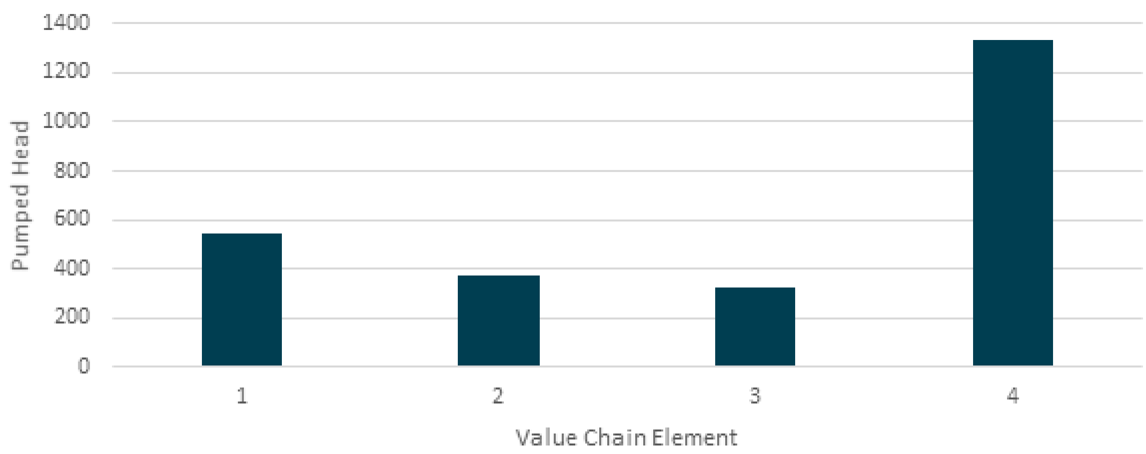


4.2 Use of disaggregated pumping head values in cost models

We recognise that pumping is a valid cost driver. However, we have concerns about the way in which pumping head is being incorporated into the draft models that have been provided. For example, Ofwat's water network plus models only includes a measure of <u>treatment</u> pumping head. Engineering logic would indicate that the water treatment process should require minimal pumping. Given this, we would expect the majority of pumping costs in the network plus control to be incurred in treated water distribution – therefore we would expect that this component of pumping head would be included in models which contain treated water distribution costs.

Furthermore, it is clear from the datashare that the majority of pumping occurs in treated water distribution.

Figure 27: Total Industry Pumped Head Across Water Value Chain in an Average Year



The table below illustrates pumping head allocations across the industry, which is averaged across the six years in the datashare. Somewhat counter-intuitively, some companies (Northumbrian, Wessex, and Welsh) report higher pumping head in treatment than in distribution. This indicates to us that there may be some material differences in calculation and reporting methodology as such a result does not have a clear engineering rationale and is counter to a priori expectations.

Company	APH: Resources	Resources confidence grade (2017)	APH: Raw water transport	Raw water transport confidence grade (2017)	APH: Treatment	Treatment confidence grade (2017)	APH: Distribution	Distribution confidence grade (2017)	Total pumping head
ANH	43.03	B4	36.41	B4	9.44	C4	73.29	B3	162.17
NES	45.59	A2	6.37	A2	32.74	A2	20.80	A2	105.50
NWT	17.55	B3	0.51	BX	4.79	B4	49.36	B3	72.21
SRN	26.98	B3	8.06	B3	14.23	C3	106.56	B3	155.82
SVT	12.77	B2	31.19	B2	15.74	B2	93.25	B2	152.95
SWT	34.31	B3	3.37	B3	6.27	B3	88.41	B3	132.35
TMS	23.05	B2	13.94	B2	16.99	B2	73.82	B2	127.80
WSH	38.27	B3	23.40	B3	57.14	B3	40.73	B3	159.53
WSX	34.62	C2	0.00	C2	77.59	C2	18.89	C2	131.10
YKY	5.07	B3	32.09	B3	8.98	B3	76.79	B3	122.92
AFW	16.86	B3	18.48	B3	14.09	C5	75.97	B2	125.39
BRL	28.58	C3	41.48	C3	8.92	C3	110.97	C3	189.95
DVW	57.00	B3	89.82	B3	19.93	B3	55.85	B3	222.61
PRT	28.06	A2	0.00	A1	1.76	A2	38.23	A2	68.05
SBW	20.32	A2	25.89	A2	7.76	A2	71.42	A2	125.38
SES	42.55	B2	20.44	B2	26.23	B2	97.91	B2	187.12
SEW	41.08	B2	2.08	B2	2.19	B2	112.59	B2	157.94
SSC	28.09	A2	16.06	A2	2.10	A2	127.63	A2	173.88

Note: this is pumping head data averaged over the six years of the datashare. Confidence grades are taken from the 2017

We consider pumping head should be included in totality or excluded. Misallocation of pumping head (or differences in reporting methodology – see section 4.3) could introduce significant distortions into cost assessment models, if a partial pumping head measure is used that is materially inconsistent between companies in a way that is not actually indicative of differences in cost. To avoid such potential distortions, the available approaches are:

- To include the aggregated total pumping head value within aggregated cost models; and
- To avoid relying on partial pumping head factors within disaggregated models

Replacing disaggregated pumping head with an appropriately aggregated measure of pumping head in Ofwat's models will place greater reliance on its engineering justification. The tables below illustrate the outcome from network plus and overall water models that do this (with the revised pumping head variable highlighted).

It's clear that in such models the more aggregated pumping head factor loses statistical significance, whilst the coefficient stays at around the same magnitude. We consider that economic and engineering rationale should take precedence over statistical significance in this instance, particularly when the alternative has the potential to introduce significant distortions into the cost assessment process. We note there is an improvement in the performance of "number of booster pumping stations per length of main" once an appropriately aggregated measure of pumping head is included. This provides further evidence that the model benefits from the inclusion of aggregated pumping head.

Therefore, we consider Ofwat should include the widest measure of pumping head across all models. If a measure has an unexpected sign, then we consider it preferable for Ofwat to omit pumping head. This would represent a better outcome than the alternative, which could unduly reward and penalise companies due to unexplained differences in splits of pumping head.

	ONP1	ONP2	ONP3	ONP4	ONP5	ONP6	ONP7	ONP8
ln(connected properties)	1.033***	1.046***	1.064***	1.084***				
	{0.000}	{0.000}	{0.000}	{0.000}				
% of water treated in complexity levels 3-6			0.004*	0.003*		0.002		0.002
			{0.050}	{0.057}		{0.201}		{0.303}
ln(weighted average density)							0.244***	0.229**
							{0.008}	{0.014}
% of mains length refurbished and relined	0.185*	0.199**	0.264***	0.265***	0.19	0.232**	0.214	0.257**
	{0.081}	{0.049}	{0.006}	{0.006}	{0.107}	{0.038}	{0.121}	{0.045}
ln(booster pumping stations per lengths of main)	0.392**	0.416***			0.450***	0.416***		
	{0.023}	{0.006}			{0.005}	{0.007}		
ln(service reservoirs and water towers per lengths of main)			0.335**	0.337***			0.253	0.237
			{0.015}	{0.009}			{0.133}	{0.153}
% of mains length laid or refurbished after 1981		-0.006		-0.006	-0.006	-0.006	-0.008	-0.009
		{0.217}		{0.158}	{0.188}	{0.123}	{0.161}	{0.108}
ln(lengths of mains)					1.041***	1.037***	1.067***	1.062***
					{0.000}	{0.000}	{0.000}	{0.000}
ln(density)					1.103***	1.064***		
					{0.000}	{0.000}		
ln(average pumping head for network plus)	**0.12**	**0.057**			**0.058**		**0.076**	
	{0.231}	{0.567}			{0.564}		{0.503}	
Constant	5.238***	5.666***	4.780***	4.774***	5.605***	5.736***	7.465***	7.733***
	{0.002}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
R squared	0.968	0.97	0.967	0.97	0.969	0.971	0.959	0.96
VIF	1.218	1.456	1.189	1.306	2.538	2.557	2.355	2.337
RESET test	0.225	0.067	0.355	0.08	0.057	0.013	0	0
N	107	107	107	107	107	107	107	107

	OWW1	OWW2	OWW3	OWW4	OWW5	OWW6	OWW7	OWW8	OWW9	OWW10	OWW11	OWW12
ln(connected properties)	1.083***	1.053***	1.101***	1.053***	1.037***	1.081***						
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}						
% of water treated in complexity levels 3-6				0.004	0.003	0.004**						
				{0.185}	{0.130}	{0.030}						
ln(weighted average density)										0.176**	0.297***	0.250***
										{0.025}	{0.000}	{0.005}
% of mains length refurbished and relined	0.2	0.202*	0.212*	0.286**	0.247**	0.276***	0.254**	0.19	0.222*	0.253*	0.17	0.217
	{0.119}	{0.054}	{0.052}	{0.014}	{0.014}	{0.006}	{0.045}	{0.109}	{0.063}	{0.066}	{0.215}	{0.118}
ln(booster pumping stations per lengths of main)		0.393***			0.392***			0.434***			0.464***	
		{0.006}			{0.006}			{0.004}			{0.000}	
ln(service reservoirs and water towers per lengths of main)			0.312***			0.336***			0.284**			0.25
			{0.007}			{0.006}			{0.022}			{0.127}
% of mains length laid or refurbished after 1981	-0.004	-0.005	-0.005	-0.005	-0.005	-0.006	-0.007	-0.004	-0.005	-0.008	-0.006	-0.007
	{0.311}	{0.261}	{0.242}	{0.101}	{0.197}	{0.178}	{0.149}	{0.259}	{0.178}	{0.144}	{0.262}	{0.150}
ln(lengths of mains)							1.098***	1.046***	1.102***	1.075***	1.006***	1.070***
							{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
ln(density)							0.755***	1.124***	1.036***			
							{0.004}	{0.000}	{0.000}			
ln(average pumping head for water wholesale)	**0.287***	**0.123**	**0.171**				**0.19**	**0.128**	**0.163**	**0.146**	**0.086**	**0.123**
	{0.083}	{0.329}	{0.216}				{0.203}	{0.320}	{0.243}	{0.344}	{0.509}	{0.400}
Constant	2.223*	5.196***	3.946***	3.696***	5.780***	4.840***	4.040**	5.093***	4.139***	6.490***	8.544***	7.191***
	{0.062}	{0.000}	{0.001}	{0.000}	{0.000}	{0.000}	{0.015}	{0.001}	{0.005}	{0.000}	{0.000}	{0.000}
R squared	0.962	0.973	0.97	0.963	0.974	0.973	0.967	0.973	0.97	0.96	0.966	0.962
VIF	1.245	1.431	1.352	1.234	1.254	1.306	1.396	2.479	2.176	1.38	3.023	2.399
RESET test	0.573	0.103	0.048	0.046	0.047	0.161	0.168	0.119	0.037	0.006	0.004	0.001
N	107	107	107	107	107	107	107	107	107	107	107	107

4.3 Inconsistencies in pumping head reporting

We are concerned that despite helpful clarification from Ofwat on pumping head methodology, inconsistencies remain in reporting which render base data unreliable. We have done some analysis using internal power costs per unit to check whether numbers in the datashare seem sensible.

Using our unit cost of pumping head, we are able to estimate how much other companies should spend on power costs, given pumping costs make up the majority of spend on power.

Company	Total Estimated Cost – based on reported pumping head (£m)	Reported power cost (£m)	Difference (£m)
ANH	35.57	28.22	7.34
NES	23.04	17.46	5.58
NWT	24.73	25.03	-0.30
SRN	16.49	12.65	3.85
SVT	55.40	42.70	12.70
SWT	11.70	10.65	1.06
TMS	65.49	50.06	15.42
WSH	25.52	19.48	6.03
WSX	8.64	7.16	1.48
YKY	30.24	25.00	5.24
AFW	22.36	20.06	2.31
BRL	10.02	7.66	2.36
DVW	2.79	1.66	1.13
PRT	2.33	1.97	0.35
SBW	3.61	2.58	1.03
SES	5.91	4.86	1.05
SEW	16.64	14.64	2.00
SSC	12.71	9.15	3.56

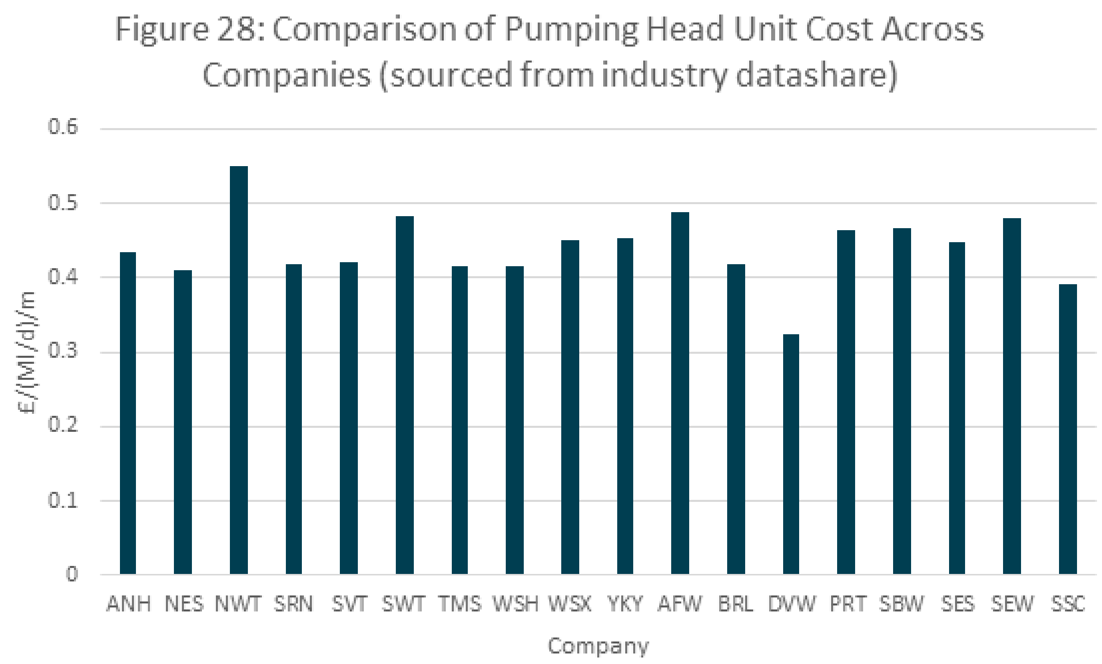
If we assume that power costs are identical across the industry (and it seems reasonable to assume that they should at least be very similar), then the variances above indicate that companies must be, variously:

1. Over-reporting distribution input;
2. Over-reporting pumping head; and/or,
3. Under-reporting power costs

It seems less likely that there would be differences in reporting methodology for power costs and distribution input – however, it is far more plausible that there is inconsistent estimation of pumping head between companies.

We can arrive at a similar conclusion using evidence from the datashare. Figure 28 shows estimated pumping unit costs across the industry. The cost was calculated as { (Power [£m]) / (Distribution input [Ml/d] x 365) / Pumped head } and averaged over the period 2012-2017 to smooth any exceptional operational spikes or falls.

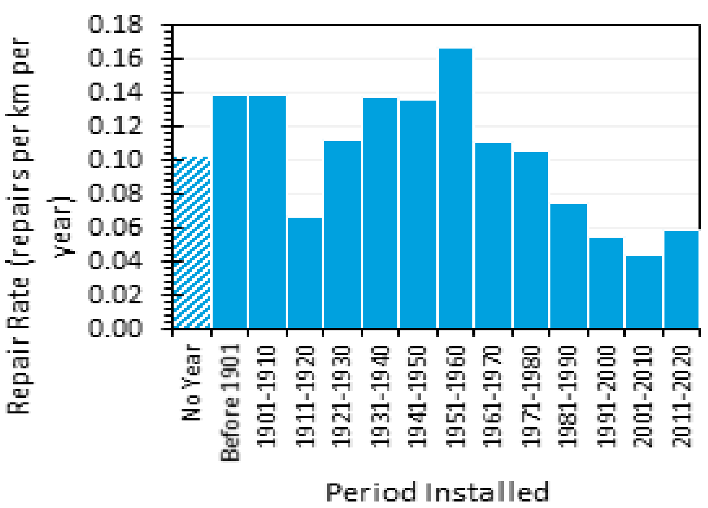
The output for this analysis appears to show that United Utilities has the highest pumping unit cost. However, such a finding is likely to be erroneous: we purchase power through a competitive national market, which has also recently been externally market tested. Again this leads to the conclusion that other companies are over-reporting pumping head. This raises questions as to the extent to which the measure is truly exogenous, or if there is significant enough uncertainty over the reporting of pumping head values to question their inclusion within cost assessment models as an explanatory variable.



4.4 Network replacement rates

We do not consider the driver used by Ofwat in water models to represent network age (percentage of mains laid or refurbished post-1981) is appropriate. Figures 29, 30a and 30b below (based on UU internal data) demonstrate that there is no set cut-off period at which cost related to network age jumps[4]. Instead, it is a rolling effect and the costs associated with certain ages of main will change in each AMP. Therefore, we do not support this measure of network age.

Figure 29 - Mains failure rates by year installed



[4] These figures are based on internal United Utilities data.

It's clear from Figure 29 that mains failure rates are high for the oldest mains before falling sharply in the period between 1910 and 1920. Following 1910, the mains failure rate rises, peaking in the period between 1930 and 1960 before falling steadily.

Figures 30a and 30b - How failure rates conditional upon mains age can be formed from individual failure rates for cohorts of mains

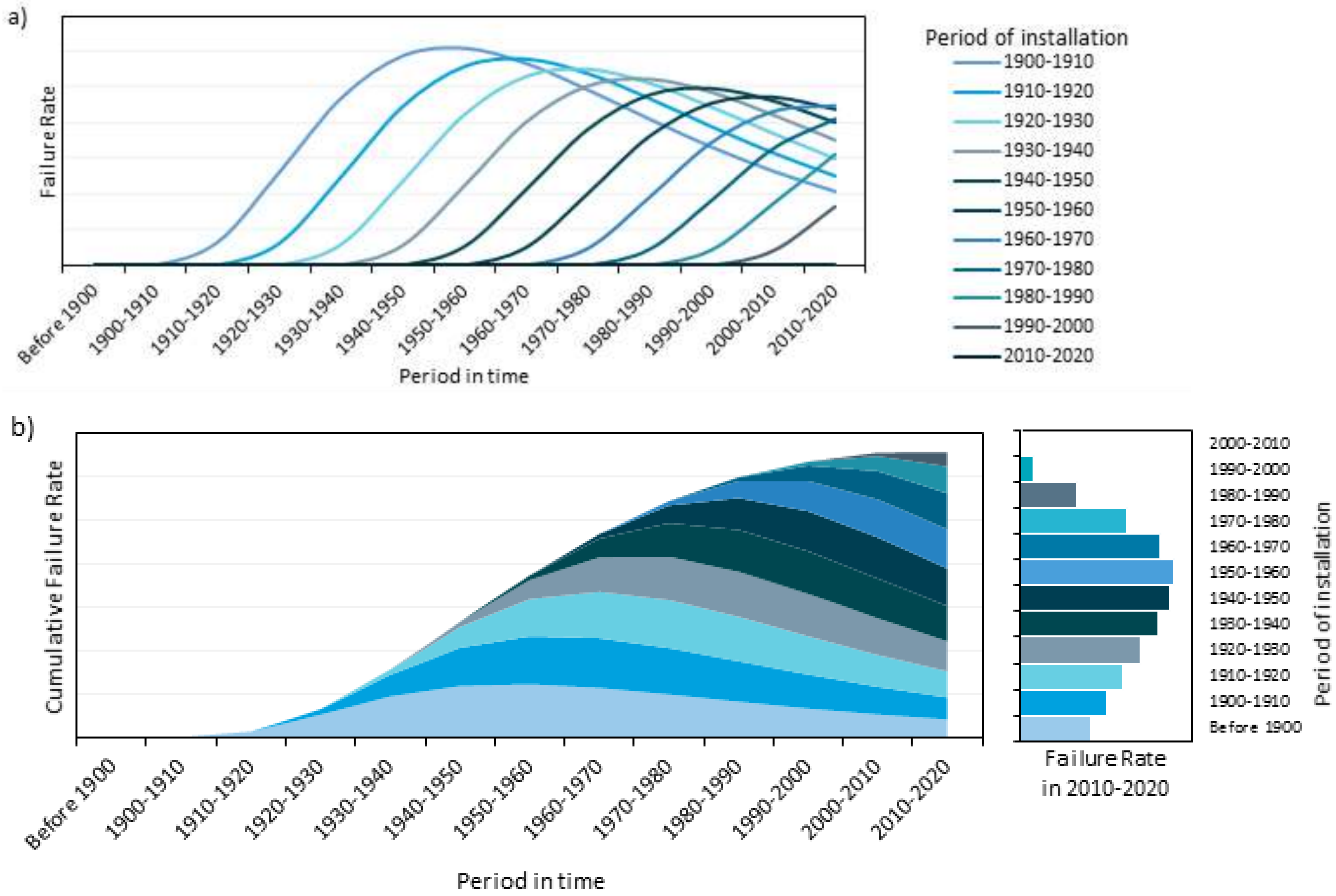


Figure 30a above shows nominal distributions of failure rates for mains laid in each decade. Each cohort follows the same lognormal distribution with a 5% improvement in failure rates each decade. Figure 30b shows the same trends as cumulative totals through the period. The proportion of failures from each period of installation will vary depending when the snapshot is taken; in the 1950s the majority of failures will be from pipe laid up until 1930 with small contributions from later mains. By the 1990s the earliest mains have passed their peak failure rate and make a smaller contribution, the majority of failing pipes will have been laid between 1920 and 1940.

The bar chart to the right shows that a snapshot taken between 2010 and 2020 would represent that mains laid in the 1950s have the highest failure rate. This would be an erroneous conclusion based on a misleading representation. The true picture revealed by the full analysis is that all mains have the same distribution of failures.

5 Comments on Retail models

5.1 Summary

Overall, we are supportive of Ofwat's approach to retail cost assessment. The consultative approach taken in the process so far has made best use of expertise from across the industry, including from United Utilities, in explaining the drivers of cost in the retail value chain. The current consultation gives industry stakeholders a further opportunity to contribute.

The use of econometric models for retail costs helps support an evidence based approach to retail cost assessment, reducing the need for a priori assumptions to be made. It also aligns well with a more symmetric approach to cost adjustment at PR19.

5.1.1 Dependent variable specification

We support Ofwat's proposed dependent variable specification. It ensures continuity from the PR14 approach of cost to serve, and reduces the chance that statistical issues such as heteroscedasticity and multicollinearity affect results. Additionally, our work with Reckon has found that this specification produces estimates with greater alignment to observed industry cost than an equivalent aggregate cost model[5].

One side effect of this specification is a low r-squared value in models of remaining retail cost. We do not consider this result to undermine the models presented by Ofwat and United Utilities. A low r-squared does not detract from the estimated relationship between cost and cost driver, and the additional precision gained in cost prediction outweighs the loss in r-squared experienced when switching from an aggregate cost model to a unit cost specification[6]**.** Therefore, we consider that a unit cost specification is better suited to econometric models that seek to predict cost.

5.1.2 Economies of scale adjustment

However, adopting a unit cost specification implicitly assumes that there are constant returns to scale across the industry.

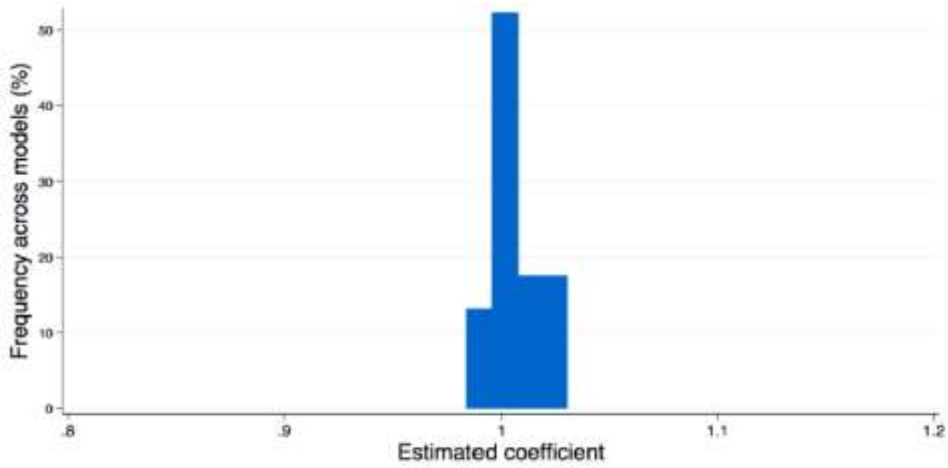
We note that management is able to take action to benefit from economies of scale in retail. These can be achieved through joint ventures and joint billing arrangements, as evidenced by actual company strategic operating decisions in the sector. We are concerned that including a measure of economies of scale would disincentivise efficient management activity, and be contrary to one of Ofwat's stated criteria in its consultation document - to minimise endogeneity within cost assessment models. We also note that regulatory focus on retail cost and incentives is relatively new. This could mean that models seeking to capture economies of scale might instead capture historical arrangements, where companies are yet to realise the opportunities for greater efficiency, which might lead to models overstating the effects of scale on efficient costs. Therefore, on balance, we consider it reasonable not to include economies of scale as a cost driver.

Our work with Reckon LLP has tested whether economies of scale exist for water and wastewater retailers. The figure below shows the histogram for the estimated coefficients on the logarithm of households that are obtained when a simple model that regresses the logarithm of remaining operating costs on the logarithm of the number of household customers is run across variations[7] to the dataset utilised by Reckon in its report.

As shown in Figure 31, the estimated coefficient on the variable relating to number of households is consistently very close to 1. This supports the use of models with no driver relating to economies of scale.

[5] Reckon LLP (2018) *Econometric models for residential retail cost assessment*
[6] Reckon LLP (2018: pg. 47) *Econometric models for residential retail cost assessment*
[7] The dataset was varied by sequentially dropping companies and years.

Figure 31 - Evidence of constant returns to scale in retail models



Additionally, the magnitude of the coefficients associated with economies of scale in Ofwat's models seem to point to scale differences across the industry that do not seem credible. The table below presents implied economies of scale from Ofwat's models when comparing two companies, one of which serves five times as many customers as the other.

Model	Coefficient on ln(number of households)	Implied economies of scale
ORDC3	-0.128	18.6%
ORDC4	-0.032	5.0%
OROC3	-0.08	12.1%
OROC4	-0.068	10.4%
ORTC4	-0.119	17.4%

Note: this table presents implied economies of scale from Ofwat's models when comparing two companies, the larger of which serves five times more customers than the smaller.

For example, Ofwat's ORTC4 model implies that unit costs would be 17.5 percent lower at the larger company. We do not consider that this ties in with economic reality, and are concerned that the inclusion of this variable could undermine Ofwat's stated aim of promoting models which are credible from an operational and economic viewpoint, given the large variation in scale across the industry.

5.2 Capturing Arrears Risk

We support Ofwat's approach in this area. We note that Ofwat has adopted one of United Utilities' measures of arrears risk, while other models test two other measures. While we consider the measures proposed by United Utilities in our submission to be the better candidates, we do recognise that the factors proposed by Ofwat also appear to be reasonable predictors of arrears risk.

We have some concerns over the use of the income measure. One potential issue with this measure is that, to our understanding, the most recent data of income deprivation for English LSOAs was published alongside the 2015 IMD, which relates to the year 2012. This seems to be a shortcoming compared to the Equifax data, which is available annually.

A further issue is that income deprivation captures financial deprivation only, whereas Equifax measures are direct measures of arrears risk. These are a more relevant cost driver for use in cost assessment models; income deprivation is undoubtedly a key driver of arrears risk, but it is not the only driver.

5.2.1 Deprivation

We note that several other companies have suggested alternative ways of capturing arrears risk. A number have adopted the Indices of Multiple Deprivation (IMD) measure. We consider this is a strong predictor of cost because it correlates well with historic deprivation and bad debt, and is calibrated to independently generated government statistics with a good track record of predicting industry arrears.

As an aside, we note the English and Welsh governments have differing methodologies for calculating the IMD. We could not be sure in the pro-forma submission that other companies have properly accounted for this difference in their datasets although we recognise the impact on the analysis is likely to be small.

However, other companies have used alternative ways of capturing arrears risk. We consider these measures have significant shortcomings and are generally unfit for purposes. We would therefore urge extreme caution when considering the potential application of these measures.

One such measure that has been suggested is unemployment. Ofwat should be aware that in practice, credit reference agencies do not use data on unemployment to assess credit risk. The fact that independent credit reference agencies reject unemployment as a measure of arrears risk should in of itself give a clear demonstration that unemployment is a poor predictor and should not be used as an explanator of differences in bad debt between companies.

Other companies have used more bespoke measures created internally. However, these companies have not shared data with us so we are unable to assess their performance as a cost driver. It is regrettable that others have not felt able to provide the same level of transparency, and therefore to invite scrutiny of their proposals for cost assessment models. Instead, we consider that the focus should be maintained on measures with a track record of predicting arrears risk and which are calibrated against widely accepted government statistics.

5.3 Bill size

We support Ofwat's use of bill size as a cost driver. We consider that its inclusion makes sense logically and operationally and note that Ofwat found it drove arrears risk at PR14. The fact that all but one company submission has included bill size as a driver suggests there is widespread support for its use at PR19.

We have found that there is a high degree of correlation between bill size and the size of the customer base that receives both water and wastewater services. This means that incorporating both factors within a model may create issues such as multicollinearity. Our work with Reckon has developed an alternative cost driver to take account of relative bill size across the industry, while removing any correlation with the dual service customer base. Results indicate that this variable performs well, even alongside a measure of dual service. We would ask Ofwat to consider adopting this measure, as it would allow one cost model to reflect a greater amount of variation in cost. More details can be found in Reckon's report[8].

5.4 Dual Service

We support Ofwat's approach to capturing the impact dual service customers have on cost. We consider it preferable to allow an econometric model to estimate a relationship, where possible, as opposed to making an ex ante assumption about that relationship.

Other companies deal with this factor by splitting the customer base into those who are single service and those who are dual service. We question whether it is optimal to take this approach in the context of a model with limited degrees of freedom. Additionally, in bad debt models proposed by Wessex and Bristol, the results suggest that a marginal cost of serving single service customers is higher than that of serving dual service customers. We fail to see any obvious logic that would underpin such a relationship and question whether this ties in with economic reality.

[8] Reckon LLP (2018: p.49) *Econometric models for residential retail cost assessment*

5.5 Metering Costs

We recognise that metering costs are a cost driver in retail. However, as set out below, we note that metering costs account for around five percent of domestic retail totex across the industry. Our assessment based on internal evidence suggests that while differences in rurality do drive costs, the difference is reasonably trivial. We also present a measure for metering which deals with how the single service/dual service customer split interacts with this factor.

5.5.1 UU evidence on metering costs

Some companies have proposed measures of metering that seek to reflect the operational cost of reading a meter, such as measures of the rural-urban split, or average traffic speed. United Utilities has sought to validate whether such factors are good reflections of what drives operational costs.

United Utilities serves some of the most densely populated areas outside of central London (the Manchester/Liverpool corridor), and some of the most rural parts of England (Cumbria). Therefore, our experience of metering costs covers operating in areas across the whole range, from the densest to the more sparsely populated areas in England and Wales, with the possible exception of the population density in central London.

Competitively tendered external contract rates reveal meter-reading costs are 12 percent higher in sparsely populated areas. Given that, on average, meter reading costs account for five percent of Domestic Retail costs, we expect that measures of the rural/urban mix would drive differences in relative cost of around 0.6 percent of total retail costs, when comparing the most urban areas and the most rural. We consider that econometric models are unlikely to reveal a cost impact of this limited scale.

The contract rates also reveal that there are no differences in Automated Meter Read (AMR) costs across different levels of density. This fact combined with other issues such as choice over the location of meters suggests that the majority of metering costs are under management control. Therefore, we do not expect the model to need to address these issues. Indeed, on a first review, we consider a majority of coefficients are biased upwards, which could reflect either overfitting or management control driving inefficient higher costs.

We would be willing to provide further evidence to support this analysis if that were of assistance.

5.5.2 Metered services

As set out above, while we do not consider that a measure of the urban/rural split is an appropriate way of dealing with meter reading costs, we recognise that meter reading does drive retail costs. Therefore, we consider it is appropriate to include a cost driver to reflect this in models of remaining retail cost.

We note that the industry and Ofwat largely use meter penetration rates or absolute customers metered. Our work with Reckon developed an alternative measure of meter penetration, called metered services provided. This variable is calculated as follows: (number of metered water only properties + number of metered sewerage only properties + 2*number of metered water and sewerage properties)/ Total number of customers.[9]

This measure recognises that where there is a different water and sewerage service provider, meter reading costs and some related costs are shared in some way between the water-only company and the water and sewerage company. Simplistic approaches to assessing meter penetration – such as those presented in some company models – do not deal with this issue.

We will illustrate this point using an example. A meter read may cost around £4[10]. If this is a dual service customer supplied by a WaSC, then this cost is used for the purposes of both water and wastewater billing. A single service customer is supplied with water services by a WOC, and wastewater services by a WaSC, with the WOC organising the meter reading. Assuming the overall cost is still £4, we would expect the WOC to recover £2 from the WaSC as the meter read is used for both water and wastewater billing. So the cost to a WaSC of a dual service meter read is double the net cost of a single service meter read. This demonstrates that the economic driver of cost is not meter penetration, but the number of metered services provided.

5.6 Transiency

We note that a number of companies have proposed transiency as a cost driver. United Utilities observes that house moves do cost money through the mechanism of opening, closing, and moving accounts. However, the cost of this is under management

[9] Reckon (2018: p.38) *Econometric models for residential retail cost assessment*

[10] This figure is used for illustrative purposes only, and doesn't accurately reflect United Utilities' cost per meter read.

control; efficient management practices such as the creation of digital interfaces allow customers to open, close and switch accounts remotely with minimal cost.

We do recognise that a household could move and leave unpaid debt behind. However, in our view, the underlying driver of debt risk is not transiency, but deprivation. To test this view, we looked at internal data on account changes and on deprivation by Lower Super Output Area (LSOA) within United Utilities' area of appointment for one year. The number of account changes per LSOA ranges from one to 474. The dataset has 4,480 observations, with a mean of 71 account changes.

To test our hypothesis, we used an OLS model with standard errors adjusted for heteroscedascity. The dependent variable was retail cost to serve in each LSOA, calculated as the total cost to serve in the LSOA, divided by the number of households in that LSOA. Two independent variables were included across different models – transiency and deprivation. The model had a log-log specification. This is because we consider there to be a multiplicative relationship between these factors, as opposed to an additive one.

Under a simple linear regression of cost to serve on transiency, the model implied that a 1% increase in transiency increases cost to serve by 0.49%, with a t score of 22.5 and r^2 of 0.13. However, when we added deprivation to the model, as measured by the average IMD score for an LSOA, the impact of transiency drops to 0.087%, with a t score of 4.84 and r^2 of 0.75 (the improved fit being due to the inclusion of deprivation, not due to transiency). This result indicates that the impact of cost to serve on transiency is negligible once deprivation is accounted for.

This leads to an important insight; transiency itself has a weak effect on retail cost.

Instead, measures of transiency seem to be picking up the effect of deprivation, such as residual unpaid bad debt. In this context, using transiency to explain retail cost does not seem optimal, given that deprivation explains a much larger portion of variation in cost to serve.

The weakness of the effect also seems to suggest why transiency is unstable in company level models of retail cost. For example, we note that Economic Insight demonstrate that 24% of transiency coefficients are negative across model suite A, while 43% are negative across model suite B[11]. We do not consider that this indicates transiency is a robust cost driver. We also note that Anglian was unable to robustly model transiency. The table below shows the implied cost of transiency across both models. Note that these workings are based on a company with United Utilities' characteristics. Monetary impacts may well change if company characteristics change.

Independent Variables Included	Implied Increase in CTS per 1% Increase in Transiency	Implied Overall Cost of Transiency	Transiency Cost as % of UU 2015/16 Total Retail Cost
Transiency	£0.19	£570,000	0.49%
Transiency and Deprivation	£0.03	£101,000	0.09%
Notes: 1) The average cost to serve per property across all UU LSOAs in 2015/16 was c.£40. 2) There were c. 320k account changes in the year. 3) UU total retail cost in 2015/16 was £116m.			

The analysis above suggests that transiency costs are immaterial and Ofwat should disregard them for inclusion in econometric models. We would be happy to discuss this issue in more detail with Ofwat.

We do recognise that central London is subject to a different challenge to the rest of the country. However, we consider that Ofwat would find it difficult to reflect this in an econometric model. Therefore, we consider that a cost adjustment factor, if justifiable, would be a more appropriate approach to reflecting this specific issue rather than a poorly targeted or reasoned adjustment to the overall cost model.

5.7 Quality of Service

We note that one company has suggested that quality of service should be included in models of other retail expenditure. We acknowledge that service levels do drive costs. However, in practice, adding a cost driver to reflect service quality to a cost

[11] Economic Insight (2018) "Population transience as a driver of household retail costs". Report for Thames Water

assessment model raises some issues. One is how this factor should be forecasted to allow it to form part of the baseline for the next price control period, and another is how customers should be protected should the forecast prove inaccurate.

We believe that existing ex-post regulatory incentives for customer service (SIM/C-MeX) are the best means of adjusting for service quality, and that no direct adjustment should be made in cost assessment models. However there is an argument that companies that consistently deliver very low quality service should be excluded from efficiency assessments given the potential for them to skew the assessment. For example a company with very poor SIM performance should not simultaneously be used to set forward looking upper quartile efficiency challenges for the industry.

5.8 Regional wages

Some companies propose the use of regional wages in retail models. We note that Ofwat has already stated it does not consider regional wages to be a cost driver.

We agree with Ofwat's position and consider costs relating to regional wages in retail to be under management control. Using this factor in cost assessment models risks rewarding inefficient management decisions.