Cathryn Ross
cathryn.ross@thameswater.co,.uk

03 February 2022

Aileen Armstrong
Senior Director
Ofwat
Centre City Tower,
7 Hill Street, Birmingham.
B5 4UA

Assessing Base Cost at PR24

Dear Aileen,

Ofwat's focus on the long-term view as a key ambition in *PR24 and beyond* is welcome in a changing environment. In common with the other water companies, Thames Water will need to invest to play our part in the transition to net zero, to respond to climate change, to maintain the resilience of our services, and to go further to protect the environment.

To make these investments there is a need for companies to be able to fund their efficient activities. The efficient assessment of base costs today is fundamental to deliver the long-term outcomes of tomorrow. It will be important, therefore, that the models that Ofwat use to determine efficient costs work in the context of a longer planning horizon.

Within this context, the PR19 models provide a good basis, although they can be improved. To put it bluntly, we are very concerned that Ofwat's models do not distinguish adequately between genuine cost efficiency and simply descoping activity. Failure to distinguish these two things, especially when combined with limited regulatory oversight of asset health and punitive incentives within Ofwat's assessment of plans for companies, whose botex falls outside Ofwat's expected tramlines, creates both the ability and the incentive for companies to cut capital maintenance rather than drive genuine efficiency. I do not mean to imply that companies have done this. But the ability and incentive to do so is a consequence of Ofwat's modelling approach. If the approach does not change, tighter price reviews (and especially with Ofwat's proposals that cost adjustment claims should be 'cost neutral') will increase the risk. We therefore welcome the inclusion of Ofwat's thinking in respect of capital maintenance in this consultation. However, we think Ofwat need to go further.

To give a more specific example of the change that is needed, we think Ofwat's benchmark models need to be adapted to deal with material, lumpy expenditure, crucially including capital maintenance. Improving this will be important for asset health and the delivery of the resilience that our customers and stakeholders need. Linked to this, but distinct from it, is the fact that the past is not necessarily a good guide to the future in terms of the need for asset maintenance. The maintenance standards we may need to adopt in order to maintain the resilience of our assets in a scenario of 2 degrees of global warming – let alone the even greater average temperatures in the South East that the Committee on Climate Change has recently set out – could be very different to those of the past. We also believe that acknowledgement is needed that in some areas past and current levels of asset replacement are simply unsustainable over the long term. For example, at the current replacement rate for sewers and water mains of 0.2% and 0.6% respectively, it will take 500 years and 167 years respectively to replace existing sewers and pipes. We set out in response to Q19 a proposal to start the process of moving to more sustainable levels of mains replacement through a separate allowance.

We welcome the overall approach of the modelling framework for the econometrics models and also the inclusion of different components related to the long-term strategy of companies. In addition to the points above we would like you to consider, in particular, the following areas that we have identified as crucial for effective cost assessment over the longer-term:

- We suggest in our response to Q1 the inclusion of environmental or climate change factors as key exogenous cost drivers, as these will have a fundamental role in the long-term planning of the industry in the next decades. In addition, it will be important that the approach to cost efficiency in the modelling does not disincentivise carbon reductions nor the development of nature-based solutions.
- The econometric base cost models need to be adjustable and updated to achieve the expected customers' demands and to maintain and improve the asset health of an ageing network across the industry. A good example of adjustable models can be found in the potential development of a robust Wholesale Wastewater Network Plus (WWWNP) econometric model. We welcome this initiative as it is important that we can capture all the synergies across the waste network between sewage collection and treatment, to maintain the sustainability of the waste network in the next decades. We have provided in this consultation response some indicative models for WWWNP that could be used as an initial starting point for the development of more robust models, built in collaboration with the industry.
- The recognition of the need to incorporate the service or quality dimension into the base cost models for PR24 assessment of efficient costs is welcome. We provide in response to question Q22 an example model that would allow leakage to be included in the botex modelling and in response to Q24 another example using a service-quality index.
- The use of CACs in both base and enhancement expenditure is an important element of any cost assessment process. We are concerned that the high evidential and materiality thresholds from PR19 already make it difficult for any claims to pass and the addition of the symmetrical adjustment approach will make it even more difficult. It is important that companies are properly funded to make the necessary investments to improve services to customers and the environment.

We are keen to work with Ofwat to help develop the benchmarking models for PR24 and beyond. My team and I look forward to working with you over the coming months. I attach our detailed response. If you have any questions, please contact Carlos Pineda at Carlos.PinedaBermudez@thameswater.co.uk.

Yours sincerely,

Cathryn Ross
Strategy and Regulatory Affairs Director

Response to Assessing Base Cost at PR24

Thames Water

03 February 2022

Response to consultation questions

2. Principles of PR24 Base Cost Assessment

Q1: Do you agree with our principles of base cost assessment?

Yes, we agree with the six principles proposed by Ofwat for PR24. However, we think that taking into consideration the aim of PR24 to focus on the long-term, they might not be sufficient as currently described for the challenges ahead.

In particular, within Ofwat's principle of focusing on exogenous cost drivers, we would invite Ofwat to consider drivers that capture **environmental or climate change** attributes. These drivers will play a fundamental role in the long-term planning of the industry in the next decades. Missing such an important dimension in the models would not be ideal in assessing the efficiency of the industry. We have included in response to Q8 an example of how rainfall could be incorporated into the modelling. Further consideration should also be given to including increasing temperatures, which for example, can result in more algal blooms, decreasing raw water quality, unless treated through additional water treatment at additional cost. Also, overall cost efficiency may not always be associated with optimal carbon efficiency. The approach to modelling needs to ensure that it doesn't create a disincentive to reduce carbon given the 2030 net zero commitment and the 2050 Government targets.

In addition, we consider that two of the principles – focus on exogenous cost drivers and sensibly simple and transparent – should not be interpreted too rigidly as this could prevent the inclusion of important elements, such as the quality or service dimension in the models. While service could be seen to be within management control and therefore excluded on endogeneity grounds, it could also be a really important omitted variable and as explained below this endogeneity could be managed.

Endogeneity is a common issue that from the *econometric* point of view can come from three different sources: i) measurement error[1], ii) omitted variables[2] or iii) simultaneity[3]. All of these

[1] Where the Industry and Ofwat make huge efforts to provide consistency in the data across different financial and non-financial variables reducing in this way the econometric issues that can be derived from measurement error.

[2] For example, when we omit a service or quality driver in the botex econometric models this could take us to an omitted variable issue that affect the consistency or unbiasedness of the estimations of all or some of the cost drivers of the model. This omission is explained by the correlation between the error term and the explanatory variables. For instance, if the service driver is leakage management and is omitted from the base econometric model of water distribution, it is likely that this service driver is highly correlated with a scale driver such as length of mains causing in this example, an omitted variable bias, see for example Destandau and Garcia (2014) "Service quality, scale economies and ownership: an econometric analysis of water supply costs". *Journal of Regulatory Economics*, 46:152–182, where it is found that: "*Issues related to service quality are crucial for water utility management and regulation. Omitting these aspects, especially when they are treated as exogenous, can lead to large biases in estimating cost functions as well as to misleading information concerning technology*" and continue "*Estimating cost function with omitted variables such as unobserved quality dimensions of output assumes that these variables are uncorrelated with variables already included in the regression. Braeutigam and Pauly (1986) estimated a cost function in the regulated automobile insurance industry with endogenous and unobserved quality and showed the existence of a bias affecting measures such as returns to scale. They also recognized that previous studies on regulated industries included quality variables in estimated cost functions but assumed the exogeneity of these variables"* In Destandau and Garcia the quality driver is considered as endogenous and observed using different indicators. They deal with the endogeneity issue through Instrumental Variables techniques.

[3] Simultaneity could be seen also with the case of cost and a quality service driver, like leakage. The cost is explained by the level of leakage and the volume of leakage is explained by the maintenance expenditures incurred to maintain

sources refer to the case when an explanatory variable is correlated with the error term of the model. It might be the case that a driver has all the sources of endogeneity or two or just one. For example, the exclusion of a service driver might be related to omitted variable bias and also to some sort of endogeneity due to simultaneity (e.g., management control). We understand that in general the management control explanation from the econometric point of view could be seen as a part of the (in)efficiency term[4] of the model that sits within the error term and that is potentially correlated with one or more explanatory variables[5]. For instance, this (in)efficiency term could be interpreted as the (in)ability of a company and its managers to make optimal decisions or to influence strategically through their decisions, certain short-term drivers, such as the case of service. This could lead to a situation of endogeneity that could cause inconsistency or biased estimations that are crucial for the calculation of efficient base cost allowances and the effect of economies of scale.

We consider that in price reviews these **endogeneity** issues could be addressed through other econometric approaches such as instrumental variables techniques, that could help to solve the correlation between explanatory variables and the error term. This would require a more flexible interpretation of the current principles.

Q2: Do you consider any important principles are missing?

In general, we agree with the approach and principles. As mentioned in our response to Q1 we consider that the list of principles may need to be interpreted flexibly. Given the aims set out in *PR24 and beyond*, in particular the long-term view, the consideration of drivers such as those related to service could have a role within the models by using the appropriate econometric techniques.[6] We believe that the principle "*Exogenous Cost Drivers* should be read in the future as "*Exogenous Cost Drivers with Treated Endogenous Drivers*". We believe that some aspects of base cost models cannot be listed as "sensibly simple", for example, the cost-service link, which would require extra analysis and effort in its estimation procedures. The extra complexity however is worthwhile so as not to close the door to opportunities that can help solve endogeneity issues and allow inclusion of important variables, such as service levels.

3. Approach to wholesale base cost modelling at PR24

Scope of wholesale modelled base costs

Q3: Do you consider the scope of wholesale modelled base costs should be amended at PR24? If so, please explain how the potential amendment/s to wholesale modelled base costs can be justified based on our proposed assessment framework.

We believe that before any attempt is made to amend the current scope or definition of base costs, we should carefully consider the econometric analysis approach within an economic regulatory framework. This will help us to understand what we are modelling. This approach

acceptable levels. Hence, these two variables are interrelated causing a simultaneous interdependence in the model. Simultaneous equation models are used in dealing with this problem, although they are challenging when implementing.

[4] For example, in the random effects model this could be the unobserved time-invariant heterogeneity random variable usually denoted by c_i or α_i in the literature.

[5] This particularly applies in a random effect or pooled models.

[6] For instance, using instrumental variables within the panel framework is an advantage to mitigate the effect of endogeneity in a potential service-quality driver included in the base models as an index. Panel datasets provide an advantage by using the lags of the endogenous variable as a potential source of instruments. These potential lagged value instruments are less likely to be influenced by current shocks. In other words, the base cost of today incurred to keep a certain level of service/improvement is likely to be unrelated with the previous levels of services.

should be based on economic principles and reflect the economic concepts that sit behind the financial/accounting data and can easily be applied to all the botex models or the more general totex approach in bioresources.

In our response to the May 2021 consultation "*PR24 and Beyond*"[7], we sought to establish the link between the accounting approach used in Ofwat's models and the economic foundations that define a cost function from the microeconomic point of view, as shown in Table 1 below. Regulated natural monopolies have a large fraction of their total costs represented as sunk capital costs. These costs are quite important both theoretically and empirically. An understanding of the cost structure and its economic implications in a regulatory framework is fundamental to assess efficiency[8].

Table 1 - Link between accounting definition and economic definition

Accounting Definition	Where is the Link?	Economic Definition	
Ofwat - General View		**Short-Run**	**Long-Run**
Totex = Opex + Capex		TC = VC + TFC	TC = VC + TFC, If Output=0, FC=0
Totex = Opex + Capex		TC = VC + (FC + SC)	
Totex = Opex + (Maintenance + Growth + Other_Enhancements)		TC = VC + (FKC + SKC + FNKC + SNKC)	
Totex = (Opex + Maintenance) + (Growth + Other_Enhancements)		TC = (VC + FKC + FNKC) + (SKC + SNKC)	
Totex = Botex + (Growth + Other_Enhancements)			
Totex = (Botex + Growth) + Other_Enhancements			
Totex = BotexPlus + Other_Enhancements			

Source: Economic Regulation, Thames Water

Where:

TC=Total Cost
VC=Variable Costs
TFC=Total Fixed Costs
FC=Fixed Costs
SC=Sunk Costs
FKC=Fixed Capital Costs
SKC=Sunk Capital Costs
FNKC=Fixed No Capital Costs
SNKC=Sunk No Capital Costs

Source: Economic Regulation, Thames Water.

We agree with the proposed criteria presented by Severn Trent (SVT) and listed in Table 3.2 of the consultation. We believe that these criteria should be understood or complemented within the economic context illustrated in Table 1 above. For example, before putting into practice the SVT criteria, we should think how the costs listed on the current definition of base costs (e.g., opex + capital maintenance + growth enhancement) are related to the different inputs of production such as Labour (L), Materials (M), Energy (E) and Capital (K), and the link they have with the different type of economic costs mentioned in the table above, such as VC, SC or FC. Understanding this, would help us to mirror these costs with the accounting definitions in the reporting tables. For instance, if the costs proposed to be included in the current scope do not fall into any of these economic cost categories (e.g., VC), we should consider excluding them from the botex definition, as this cost will not provide any economic link that is fundamental to the assessment of efficient costs within a regulated industry. This economic approach would also help to avoid econometric inconsistencies in the modelling process.

Moreover, if the costs can be related to the inputs of production, as described above, we could link the proposed cost to be included in the definition as a way to understand the potential degree of cost complementarities among the current costs or inputs of production by calculating proportional costs among these inputs.

This accounting and economic link could also be beneficial in the determination of the level that we should take into modelling purposes of the costs. For instance, starting at the totex level we can see the economic and econometric implications that this level of costs can put in the

[7] See answer to Q10.1 in Thames Water Response.

[8] See Joskow, P. (2007). Regulation of Natural Monopoly. *Handbook of Law and Economics*, Vol. 2. p.1241. Elseiver.

specification of the models. At this level we would need to think carefully on how to deal with variable costs and fixed or sunk costs and the econometric and economic implications of taking this potential approach to get sensible econometric results[9].

We also believe that the time distinction between short v long run cost functions should be considered in the scope as this would be in line with Ofwat's aim of including a long-term focus in PR24. This can be crucial for a consistent and stable expenditure of capital maintenance in the next decades, for example. The recognition of this could provide insights on what parts of the costs are being separated in the current modelling approach, which could help us to understand why certain types of costs might not be fully covered by the short-term five-year allowances provided by the models.

On the specific issue of separate modelling for growth expenditure we agree that this should be explored. There are some issues that will need consideration to develop good growth models, in particular:

- Growth rates are affected by a number of factors including government policy and incentives and historical growth rates and costs may not be representative of future costs.
- Cost drivers of growth in terms of "ONS household numbers" and "population" do not account entirely for the growth costs incurred in infrastructure and non-infrastructure requirements. Other factors include concentration of growth (focussed on small areas with significant expenditure requirements as opposed to growth spread evenly everywhere), movement of population within an area, and the lumpiness of non-infrastructure investment e.g., a new storage reservoir or WTW upgrade is a binary decision point not a smooth linear expenditure.

This area is suitable for use of dynamic panel models to deal with lumpiness as previously raised with Ofwat.

Q4: Would you recommend collecting additional data in relation to growth expenditure (cost and/or cost driver data) to improve cost assessment at PR24? If so, what additional data would you recommend collecting? Please provide definitions alongside suggested data additions.

The ANH suggestion for additional wastewater growth data collection seems reasonable, and this should also have a similar equivalent in water. We are not entirely convinced how this would feed a companywide econometric model as this is driving towards investment triggers for catchment specific upgrades, which feels like an enhancement case (albeit there could be a common methodology and process for it).

The growth data improvements that have been made in the latest APR are only reflected from 2020-21 data. We believe that we could recast data back to 2015-16 with some reasonable assumptions (most of this was required as part of the PR19 data tables).

If growth costs are modelled as a separate model, it is likely that this type of enhancement expenditure will have some degree of lumpiness. If that is the case and as we have mentioned in

[9] A good example on how this link should be used can be find at Destandau and Garcia (2014) "Service quality, scale economies and ownership: an econometric analysis of water supply costs". *Journal of Regulatory Economics*, 46:152–182. This paper estimates a Botex model (e.g., Opex + Maintenance) which is then treated as a Variable Cost Function within the Short-Run framework.

our paper published in the Future Ideas Lab, we believe that dynamic models would help to obtain a more objective assessment of growth expenditure[10].

Sample period selection

Q5: Do you agree that we should utilise the full historical data series available to develop the wholesale base cost models at PR24 (from 2011-12 onwards) unless there is clear justification for using a reduced time series (e.g., structural break that cannot be addressed through other remedies)?

In principle it is sensible to use more observations if the data adds new information to the model. However, it should be recognised that more observations do not necessarily mean better models. We think that it is important to clarify that in an industry with the characteristics of water and especially waste, where the within company variation of most of the drivers in the cost models is low (for example length of mains does not change materially from year to year)[11] it is not just the number of observations[12] that matters in the estimations of these static panel models. In addition, the variation, year by year, within each company is important for good estimation procedures and performance of the models. Adding more of the same[13] would not necessarily help to improve the models given the **small sample** size of the industry i.e., the cross-sectional dimension of the panel dataset, where in water we have 17 companies and in waste only 10. These characteristics of the panel implies a very limited scope of exploiting the models to explain the variance in costs across the industry as the degrees of freedom are quite restricted.

We believe that it is also important to understand the dimensions of the panel before we take any decision on how long the sample period should be. For consistency[14] of the estimators in traditional panel data models, the theory relies on asymptotic properties or what is usually called **short panels,** which does not mean **small samples**. By short panels we should understand that the cross-sectional dimension of the panel is defined by a large number of N (or a large number of companies) and it is usually defined as $N \rightarrow \infty$ or where N tends to infinity, which in other words means that for consistency of the estimators of our models, we require a **large sample** in the N or cross-section dimension of the panel, which is clearly not the case in the industry where N tends to be small and fixed. The other dimension of the panel represented by the time or period is usually represented by T which under a short-panel structure is considered as fixed or small when compared to N. Under these short-panel characteristics, the models provide enough confidence to use cluster-robust standard errors that require that $N \rightarrow \infty$. This could be an issue in panels that are based on regional data, like the water or waste industries, that is basically fixed in N and only increases over the time T dimension, or what is commonly known **long panels** where T>N. This clustering is relevant when we are investigating the statistical effect of a cost driver. Last but not least, this short-panel dataset can also provide enough confidence for the consistency of the estimators, which is crucial in the determination of costs allowances, whereas in a long-panel the circumstances can be different.

[10] See paper at https://www.ofwat.gov.uk/wp-content/uploads/2021/10/TMS-Dynamic-Panel-Data-Modelling-Approach-for-Enhancements.pdf

[11] This mean that the dynamics over time of most of the cost drivers barely change.

[12] Where the number of observations is the total number of companies (N) in the model times the years (T) they are observed, NxT.

[13] See some of the issues brought by small samples in the more general framework of Micronumerosity or small sample size in Goldberger (1991), A Course in Econometrics, Harvard University Press, p. 248.

[14] For the **consistency** of an estimator, we should understand the framework when it gets closer and closer to its true value the larger the sample is, where in traditional panel data models this refers to the N dimension or number of companies in the industry. It is also important to realise that when we talk about **unbiased** estimators, we are referring to finite sample properties that is not affected by increasing the sample size, either in N or T dimensions.

When we refer to the small sample in our dataset, we are pointing specifically to the N dimension of the panel that is not only fixed due to the regional characteristics of the industry but also it is small in size, in particular the wastewater industry. In other words, this is what defines the **small sample** properties of the industry. These characteristics of the N dimension (regional fixed and small in size) takes us to a situation where the small sample of the industry could jeopardize the consistency of the estimations even more within a **long-panel** data framework that could bring time series data issues such as the structural break example mentioned in the consultation[15].

In this context, if we estimate a model based on data from 2011-12 to 2023-24 for a static panel model, the T dimension of the panel will be based on 13 years, which would be greater than the number of companies in the waste industry N=10, putting in doubt some consistency properties of the estimators[16]. For example, if the RE model is estimated by feasible GLS (FGLS) where T is large, and N is small, the efficiency of the estimator would disappear[17]. Therefore, in water we think that the panel dimensions would be in a better position and significant changes might not be needed for both dimensions of the panel, N and T, whereas in waste we suggest using a time dimension that covers a maximum of nine or ten years for a historical panel data analysis. More than that should be considered with caution.

We need to note that the T time dimension could be increased further if forecast data from business plans are included for a full waste and water analysis, yielding around 18 years of time. In this scenario the waste panel dataset should be considered carefully whereas in water the dimensions might still hold but by a narrow margin. A pragmatic alternative is to move the panel in the waste case, for example from 2015-16/2016-17 to 2023-24 which translates into 8 or 9 years of historical data coming back to the N>T panel data standard framework for consistency of the estimators and robustness of the cluster-standard errors. If forecast data is included similar adjustments should be considered in the analysis.

Q6: Should we consider including business plan forecasts in our wholesale base cost models at PR24?

We believe that including business plans (BP) in the base cost models is, in theory, a pragmatic way to incorporate the long-term effects. However, Ofwat have created strong incentives on companies to ensure its business plan projections do not exceed what Ofwat 's models are likely to find as an efficient level. This acts a significant barrier to companies proposing significant increases. The business plan forecasts will also reflect other factors including uncertainties and unachievable innovation or technologies that could not be realistic in the mid or long-term periods considered. The data in business plans is therefore likely to be a combination of efficiency and risk and could bring some uncertain scenarios. In other words, the business plan data might not reflect efficient business plans, which could significantly distort the picture of what a forward-

[15] In this case, when T is not much larger than N, the results of the model can face finite-sample bias in its estimators and standard errors. In this scenario where T>N it is suggested to use **xtregar** and use the different alternatives proposed by the model in the case of random effects. This model incorporates an auto-regressive or AR(1) error $u_{it} = u_{i,t-1} + \varepsilon_{it}$ which is a better approach than the i.i.d. error assumed under **xtreg** that is $u_{it} = \varepsilon_{it}$. In this way this potentially approach can bring more efficient parameter estimates. For a practical discussion on these issues see Cameron and Trivedi (2010), Microeconometrics Using Stata.

[16] See Wooldridge (2010). Econometric Analysis of Cross Section and Panel Data. *MIT, press*, second edition. p. 285.

[17] See Schmidt and Sickles (1984). Production Frontiers and Panel Data, *Journal of Business & Economic Statistics*, p. 369. For more details on the effects of having a small sample into GLS methods and unbiases of the estimator or the estimation of the variance parameters of the composite error, see Ullah (2004). Finite Sample Econometrics, *Oxford University Press*. p. 123.

looking catch-up efficiency challenge should look like. If this is implemented, Ofwat will need to be careful on selecting the most "objective" forward-looking frontier, otherwise this could have significant, and material impacts driven by uncertainty-risks and not by genuine efficiency.

If forecast data is used, for the period 2024-25 to 2029-30 covering 6 years of the T time dimension of the panel, one practical way that could be explored is ***to use separate assessment of the sample periods*** using the same structure of the models between historical and forecast data. An efficiency catch-up challenge can then be calculated using both approaches and any difference considered. If this difference is significant it could be an indication of the extra noise that is brought by the forward-looking data. We would expect more weight to be given to the catch-up challenge based on the historical data given the potential for noise to be included in the forecast business planning data. We consider that it is important that separate assessment is made with and without the forecast business plan data.

Target modelling suite

Q7: Do you agree with our proposed target wholesale base cost modelling suite at PR24?

We support the continuation of the different sets of aggregate and disaggregated models at PR24 and the triangulation of these different set of models. Moreover, we welcome the initiative of developing a potential wastewater network plus model for PR24 given the changing circumstances of taking bioresources out of the wastewater assessment.[18] Although, we recognise the challenge to develop a robust model for the wastewater network given the small sample size and its effects on degrees of freedom and explanatory variables in the model.

For wholesale water, if an additional robust Treated Water Distribution (TWD) model is developed this could complement the current assessment of this part of the value chain that currently has only one model. Table 2 below shows all the results of the models used by the CMA. We have noticed that the current TWD model covering the period 2011-12 to 2019-20 used by the CMA presents a drop in the p-value of the functional form using the RESET test when assessed at the 10% significant level and compared to the PR19 Ofwat period of assessment. This also is the case for the wholesale water model 2 (WW2).

[18] See answer to question Q8 where we propose some wastewater network plus models.

Table 2 – Water CMA Models

Static Panel-Data Models: Wholesale Water CMA Models (Sample Time Period: 2011-12 to 2019-20)

	WRP1	WRP2	TWD1	WW1	WW2
	b/se	b/se	b/se	b/se	b/se
Ln(Properties)	1.033***	1.030***		1.036***	1.024***
	(0.046)	(0.057)		(0.031)	(0.030)
Complexity36	0.008***			0.006***	
	(0.001)			(0.001)	
Ln(Weighted_Density)	-1.451***	-0.958**	-3.338***	-2.371***	-1.939***
	(0.451)	(0.423)	(0.447)	(0.348)	(0.262)
Ln(Weighted_Dsty)^2	0.091***	0.055*	0.266***	0.168***	0.137***
	(0.031)	(0.030)	(0.031)	(0.024)	(0.018)
Ln(Weighted_Comple~)		0.444***			0.533***
		(0.160)			(0.115)
Lenghts of main			1.055***		
			(0.040)		
Number of booster ~o			0.570***	0.316***	0.324***
			(0.150)	(0.111)	(0.098)
constant	-5.307***	-6.979***	6.782***	-0.331	-1.948*
	(1.600)	(1.437)	(1.510)	(1.309)	(1.008)
R2_Overall	0.929	0.916	0.962	0.973	0.975
Wald_Chi2	1698.334	1099.808	1745.518	3239.399	3170.217
RESET_P_value	0.49	0.19	0.05	0.16	0.06
BPagan_Test_P_value	0.00	0.00	0.00	0.00	0.00
Serial_Corr_P_Value	0.01	0.00	0.00	0.00	0.00
StdDev_Ind_Effect	0.20	0.21	0.17	0.12	0.11
StdDev_Idiosy_error	0.20	0.20	0.14	0.13	0.13
Corr_comp_error	0.51	0.55	0.62	0.45	0.39
N_Sample_Size	21.00	21.00	21.00	21.00	21.00
T_Sample_Size	4.71	4.71	4.71	4.71	4.71
Observations	158.00	158.00	158.00	158.00	158.00

Source: Economic Regulation, Thames Water. Note: T_Sample_Size represents the unbalanced panel.
* p<0.10, ** p<0.05, *** p<0.01

The RESET test suggests a potential misspecification of the functional form for models TWD1 and WW2 at the 10% level, although at the 5% level the test would indicate that no issues on the functional form are present. We invite Ofwat to clarify its approach to this type of statistical outcome and interpretation of the results. For example, is the 5% and 10% levels of significance both acceptable in to Ofwat in its consideration of an acceptable functional form? We believe that flexibility between 5% and 10% significance should be accepted to assess the functional form of the models.

Table 2 also shows other statistics related to the presence of an unobserved time-invariant heterogeneity random variable[19] as suggested by the results shown by the Breusch-Pagan test p-value across all models. In addition, we have added some information regarding the standard deviation of this time-invariant unobserved heterogeneity and the idiosyncratic error term, suggesting that the former effect is more important within the water distribution model (e.g., 0.17 > 0.14). This can also be seen through the interclass correlation of the composite error term, 0.62 which can be seen as the ratio of the variance of the unobserved heterogeneity and the composite error which is useful to see the relative importance of this unobserved effect, providing more information for the use of random effects models.

[19] This unobserved heterogeneity effect is specific for each company and it can be called unobserved firm heterogeneity. It represents all factors affecting companies' costs that do not change over time. Management skills could be captured in this effect or any other geographical characteristic of the regions where companies operate that are constant across time, for example.

We also consider that for water resources plus, analysis should be undertaken on cost drivers related to the water resources part only, which could be added in the models, such as the sources where water is extracted from, rivers, boreholes etc. We would like to provide in the near future some models regarding this. Due to time constrains it was not possible to develop such a model for this consultation response.

Q8: Do you consider it would be worthwhile attempting to develop wholesale wastewater network plus models for PR24? If so, do you propose any potential wastewater network plus cost model specifications to consider?

In principle, we support the attempt to develop a wholesale wastewater network plus model for PR24 but given the challenges of the sample size and its effects on the list of cost drivers in the model, we believe that different pragmatic solutions would be necessary to mitigate the impact of the challenges, for example the use of triangulation between bottom-up and top-down models.

In principle, we believe that a model that captures the network characteristics of the natural monopoly in wastewater is important when assessing efficiency. For example, this approach could capture better the different interdependencies derived between collection and sewage treatment business units that are reflected in cost efficiency complementarities across the network and the synergies that this interaction could generate in incentivising efficiency. Nevertheless, we also need to recognise the potential challenges driven by the small sample size and the limited scope of including relevant cost drivers in the models.

Within this context, the choice of a few very good cost drivers is imperative to have a robust model and so we have been exploring some potential models for wholesale wastewater Network Plus (WWWNP) that could be used as a starting point to develop future models for PR24. We start the analysis by investigating the three main scale drivers available that describe the size of the wastewater network. Table 3 below "*Scale Drivers*" show the results when the scale driver is represented by Load, Sewer Length and Properties, separately.

Table 3 – Scale Drivers

Scale Drivers: Waste Network Plus Econometric Models (Sample: 2011-12 to 2019-20)

	Scale1 b/se	Scale2 b/se	Scale3 b/se
Ln(Load)	0.743*** (0.057)		
Ln(Sewer_Length)		0.833*** (0.092)	
Ln(Properties)			0.802*** (0.072)
constant	-3.854*** (0.733)	-3.436*** (1.009)	-6.104*** (1.051)
R2_Overall	0.919	0.856	0.916
Wald_Chi2	167.562	81.879	124.846
RESET_P_value	0.04	0.23	0.01
BPagan_Test_P_value	0.00	0.00	0.00
Serial_Corr_P_Value	0.00	0.00	0.00
StdDev_Ind_Effect	0.12	0.18	0.13
StdDev_Idiosy_error	0.09	0.09	0.08
Corr_comp_error	0.66	0.82	0.71
N_Sample_Size	10.00	10.00	10.00
T_Sample_Size	9.00	9.00	9.00
Observations	90.00	90.00	90.00

Source: Economic Regulation, Thames Water.
* p<0.10, ** p<0.05, *** p<0.01

Interestingly, the scale driver on its own explains between 85%-92% of the variance of botex cost across the industry. This shows how powerful is the scale driver in the models. Among the three scale drivers the load provides the highest R^2 and length of sewer the lowest. Table 4 "*Scale Driver Models Extended*" shows the econometric results when the scale models 1,2 and 3 are extended with two more relevant drivers for sewage network.

Table 4 – Scale Driver Models Extended

Scale Drivers Models Extended: Waste Network Plus Econometric Models (Sample: 2011-12 to 2019-20)

	Scale1_B13 b/se	Scale1_B6 b/se	Scale2_B13 b/se	Scale2_B6 b/se	Scale3_B13 b/se	Scale3_B6 b/se
Ln(Load)	0.684*** (0.058)	0.708*** (0.060)				
Prp_NH3_Below3	0.004*** (0.000)	0.004*** (0.001)	0.004*** (0.000)	0.004*** (0.000)	0.005*** (0.001)	0.005*** (0.001)
Prp_Band13	0.026* (0.015)		0.032 (0.020)		0.013 (0.028)	
Prp_Band6		-0.010** (0.004)		-0.011* (0.006)		-0.006 (0.007)
Ln(Properties)			0.773*** (0.101)	0.785*** (0.082)		
Ln(Sewer_Length)					0.699*** (0.128)	0.713*** (0.102)
constant	-3.304*** (0.794)	-2.735*** (0.672)	-5.899*** (1.528)	-5.127*** (0.984)	-2.161 (1.496)	-1.826** (0.838)
R2_Overall	0.917	0.903	0.916	0.899	0.878	0.862
Wald_Chi2	322.931	349.540	356.455	403.433	255.551	325.123
RESET_P_value	0.54	0.52	0.23	0.34	0.22	0.17
BPagan_Test_P_value	0.00	0.00	0.00	0.00	0.00	0.00
Serial_Corr_P_Value	0.00	0.02	0.01	0.06	0.01	0.03
StdDev_Ind_Effect	0.13	0.14	0.14	0.15	0.20	0.19
StdDev_Idiosy_error	0.08	0.08	0.08	0.07	0.08	0.08
Corr_comp_error	0.73	0.76	0.76	0.80	0.86	0.86
N_Sample_Size	10.00	10.00	10.00	10.00	10.00	10.00
T_Sample_Size	9.00	9.00	9.00	9.00	9.00	9.00
Observations	90.00	90.00	90.00	90.00	90.00	90.00

Source: Economic Regulation, Thames Water.
* p<0.10, ** p<0.05, *** p<0.01

Based on the results of the previous table, we notice that the models with length of sewer as a scale driver is the one with the lowest R^2 and where the size bands 1-3 and 6 were not statistically significant. Given the multicollinearity issues that these scale drivers present, if added all together in a single model specification, we decide to build the wastewater network model from the *load* scale driver. If a second scale driver is also considered as an alternative to the load driver, we consider that the number of properties would provide more explanation about the variability of costs across the industry when compared to sewer length. We do not see sewer length as a strong scale driver when attempting a model for the wastewater network. This may be due, in part, to the potential for measurement error as there are substantial unmapped sewers arising from the transfer of shared sewers. This is a *similar* characteristic found in the wholesale water models where number of properties is used as the main scale cost driver.

The next table below "*Potential Botex + PR24 Waste Network Models*" presents some models using load as the scale main driver with some of the cost drivers in sewage collection and treatment used in the PR19 models. As previously mentioned in response to question Q5, the small sample characteristics of the waste industry is particularly restrictive when compared to water. This restriction could impose some challenges in the modelling process of a network plus model that could be reflected in the degrees of freedom of incorporating different drivers in a WWWNP model. We think that estimating no more than five or maximum six parameters including the constant is difficult when the degrees of freedom of the model are limited and the potential overfitting that could be created by adding more drivers in such conditions[20]. This is an important characteristic to be considered in developing models in general in wastewater, either at the network plus or bioresources level. In fact, at PR19 this was the pattern found in all the models[21].

[20] It is not just about the number of observations NxT that counts for the calculation of degrees of freedom (dof) but also its variation. Adding more of the same do not necessary extend the dof and given the small characteristics of the sample, and by small we mean the fixed dimension of N or number of companies, adding more observations across time T dimension is not always the answer, especially if most of the drivers present a small within variation. Hence, we think that the dof seen as N-k is very limited, where k is the number of parameters in the model.

[21] Although the CMA has a model with five parameters in sewage collection when the square of density is included using weighted population density but not when using average property density, which could be seen as inconsistent in definition.

Table 5 – Potential Botex + PR24 Waste Network Models

Static Panel-Data Models: Potential PR24 Botex+ Waste Network Plus Econometric Models (Sample: 2011-12 to 2019-20)

	WWWNP1 b/se	WWWNP2 b/se	WWWNP3 b/se	WWWNP4 b/se	WWWNP5 b/se	WWWNP6 b/se
Ln(Load)	0.684***	0.708***	0.792***	0.830***	0.781***	0.783***
	(0.058)	(0.060)	(0.051)	(0.062)	(0.077)	(0.086)
Prp_NH3_Below3	0.004***	0.004***				
	(0.000)	(0.001)				
Prp_Band13	0.026*		0.036**		0.038**	
	(0.015)		(0.015)		(0.017)	
Prp_Band6		-0.010**		-0.017***		-0.016***
		(0.004)		(0.006)		(0.006)
Ln(Property_Density)			0.742*	1.202**		
			(0.420)	(0.562)		
Ln(Weighted_Density)					0.143**	0.248**
					(0.072)	(0.120)
constant	-3.304***	-2.735***	-7.401***	-8.128***	-5.515***	-4.936***
	(0.794)	(0.672)	(1.527)	(1.596)	(0.876)	(0.729)
R2_Overall	0.917	0.903	0.932	0.916	0.916	0.883
Wald_Chi2	322.931	349.540	356.894	283.342	278.162	263.688
RESET_P_value	0.54	0.52	0.08	0.32	0.03	0.06
BPagan_Test_P_value	0.00	0.00	0.00	0.00	0.00	0.00
Serial_Corr_P_Value	0.00	0.02	0.00	0.02	0.01	0.05
StdDev_Ind_Effect	0.13	0.14	0.12	0.12	0.13	0.14
StdDev_Idiosy_error	0.08	0.08	0.08	0.08	0.08	0.07
Corr_comp_error	0.73	0.76	0.67	0.72	0.75	0.79
N_Sample_Size	10.00	10.00	10.00	10.00	10.00	10.00
T_Sample_Size	9.00	9.00	9.00	9.00	9.00	9.00
Observations	90.00	90.00	90.00	90.00	90.00	90.00

Source: Economic Regulation, Thames Water.
* p<0.10, ** p<0.05, *** p<0.01

Within this context, the econometric results in Table 5 above "*Potential Botex + PR24 Waste Network Models*" start with the PR19 specification of sewage treatment models but are extended to cover botex expenditure of the network in the period 2011-12 to 2019-20. Looking at models WWWNP1 and WWWNP2 in table 5[22], the statistical results of these two models show a consistent high R^2 around 0.90, a high p-value for the functional form specification test[23], a strong presence of the unobserved heterogeneity effect[24] and all the drivers are statistically significant to some degree. The next four models in Table 5 (WWWNP3 to WWWNP6) are the same as mentioned before but replaced by the complexity or quality of the treatment works drivers and by two different measures of density, weighted population density and average property density. When density is introduced instead of complexity of the treatment, the results remain quite robust for the scale load driver. This result shows a robust presence of economies of scale through the magnitude of the Load coefficient being below one. The effect of density seems to be statistically significant in all models WWWNP3 to WWWNP6, although in some specifications the functional form test fails (see model WWWNP5). We tried some square effects of density, but the models did not perform well overall plus the square term and functional form where not robust when this factor is added.

[22] These models are the same as the one labelled Scael1_B13 and Scale1_B6. Presented in table 4 "*Scale Drivers Models Extended*".
[23] This was not the case when the model only treats Load as the only variable of the model, see the table Scale Drivers.
[24] See in the table results for example the Breuch-Pagan test p-value for RE model v Pooled OLS and the standard deviation of the individual effect over the idiosyncratic error term (StdDev_Ind_Effect) or the Correlation of the composite error term).

The set of models presented to this point could be used in principle as a starting point in developing other models that could extend the list of drivers to some degree, considering the limiting degrees of freedom we have *N-K*, where *K* is the number of parameters and *N* the number of companies[25].

Table 6 "*Waste Network Plus models with Density*" shows the results of adding density and complexity in the models at the same time. They show that the density effect suggested by the previous models (WWWNP3-6) disappears when assessed with complexity and the other cost drivers load, band1-3 and band6. The band1-3 driver effects vanish with the inclusion of density.

Table 6 – Waste Network Plus models with Density

Waste Network Plus Econometric Models with Density (Sample: 2011-12 to 2019-20)

	WWWNP7 b/se	WWWNP8 b/se	WWWNP9 b/se	WWWNP10 b/se
Ln(Load)	0.692***	0.665***	0.734***	0.685***
	(0.062)	(0.075)	(0.070)	(0.095)
Prp_NH3_Below3	0.003***	0.004***	0.003***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
Prp_Band13	0.032*	0.030		
	(0.018)	(0.021)		
Ln(Property_Density)	0.463		0.884	
	(0.547)		(0.676)	
Ln(Weighted_Density)		0.051		0.161
		(0.112)		(0.144)
Prp_Band6			-0.016**	-0.015**
			(0.006)	(0.006)
constant	-5.148**	-3.440***	-5.911**	-3.183***
	(2.266)	(0.960)	(2.348)	(0.838)
R2_Overall	0.923	0.913	0.911	0.880
Wald_Chi2	476.243	267.700	734.569	233.353
RESET_P_value	0.31	0.12	0.37	0.06
BPagan_Test_P_value	0.00	0.00	0.00	0.00
Serial_Corr_P_Value	0.01	0.01	0.03	0.05
StdDev_Ind_Effect	0.12	0.14	0.12	0.16
StdDev_Idiosy_error	0.08	0.08	0.08	0.07
Corr_comp_error	0.68	0.78	0.73	0.82
N_Sample_Size	10.00	10.00	10.00	10.00
T_Sample_Size	9.00	9.00	9.00	9.00
Observations	90.00	90.00	90.00	90.00

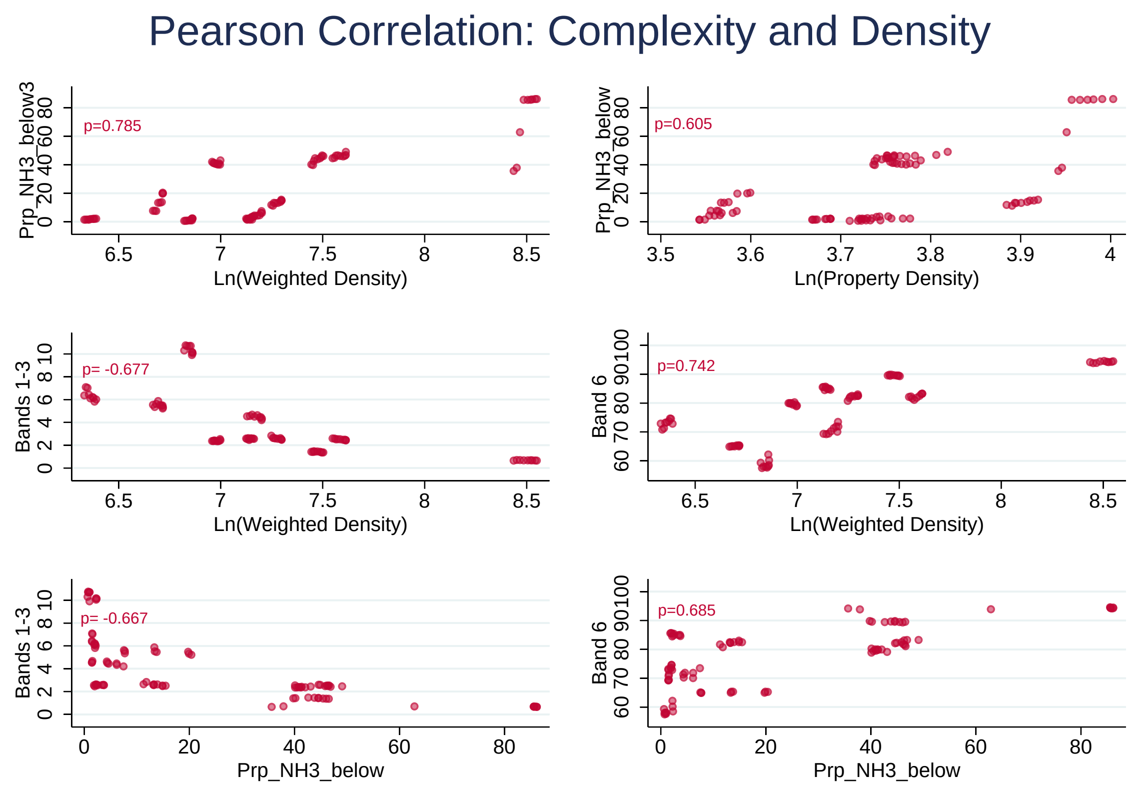
Source: Economic Regulation, Thames Water.
* p<0.10, ** p<0.05, *** p<0.01

A potential explanation of why density is not having a significant effect in these specifications is due to the negative correlation between density and the load treated in band size 1-3 and band 6. Basically, the denser a region, the more load is treated at higher bands or in other words the higher the density the less load is treated at bands 1-3. In addition, the percentage of load NH3 below 3 mlg has a positive correlation the denser the area is (e.g., weighted population density, average property density or load treated at high bands). This could indicate that the variables complexity and density are capturing similar things at the wastewater network plus model level, which could indicate that using one of the two drivers is sensible for a WWWNP model. We believe that testing these two drivers in different specifications would provide better information when choosing a driver. We think that complexity of the treatment performs robustly compared to the case when density is the driver. Similar patterns are also found with the ammonia levels of

[25] Please notice that we do not consider the T dimension in this calculation as most of the cost drivers have a very low degree of within variation as explained in the answer to question Q5.

complexity and the different proportions of load treated at the different band sizes. The charts presented below in Figure 1 show the Pearson's correlation of these complexity and density drivers as an illustration of the previous description:

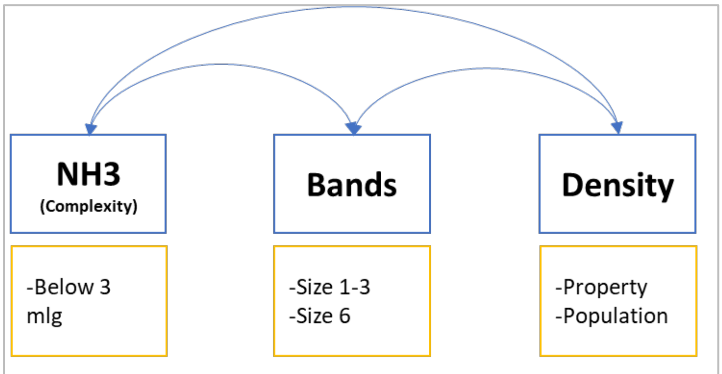
Figure 1 – Correlation: Complexity and Density



Source: Economic Regulation, Thames Water.

These facts could help us to understand what degree we are able to ask the models to control for using these different drivers that might be capturing similar effects at the network plus model level. Under this logic we have tested all the different model specifications combinations combining separately each of these complexity, density and band sizes drivers, similarly as the ones presented in Table 5 "*Potential Botex + PR24 Waste Network Models*". The diagram and the arrows below in Figure 2 illustrate the different combination paths of specifications we explore:

Figure 2 – Combination of Models



Source: Economic Regulation, Thames Water.

The econometric results based on the above diagram are shown in Table 7 below "*Complexity, Density and Band*". For example, the first column shows the combination of *Complexity and Density*. The first column shows the model with a scale load driver and the complexity of ammonia and one of the density drivers (see *Compl_Dsty1*) whereas column two is the same as the previous model described but using the other density driver (weighted population density). Similarly, Table 7 goes on with all the different combinations of *Density and Bands*, or *Complexity and Bands*. We show all these results to support our conclusions.

Table 7 – Complexity, Density and Band

Starting Point for Waste Network Plus Econometric Models: Complexity(Compl), Density(Dsty) and Band (Sample: 2011-12 to 2019-20)

	Compl_Dsty1 b/se	Compl_Dsty2 b/se	Dsty1_Band13 b/se	Dsty2_Band13 b/se	Dsty1_Band6 b/se	Dsty2_Band6 b/se	Compl_Band13 b/se	Compl_Band6 b/se
Ln(Load)	0.586***	0.592***	0.792***	0.781***	0.830***	0.783***	0.684***	0.708***
	(0.050)	(0.087)	(0.051)	(0.077)	(0.062)	(0.086)	(0.058)	(0.060)
Prp_NH3_Below3	0.003***	0.004***					0.004***	0.004***
	(0.001)	(0.001)					(0.000)	(0.001)
Ln(Property_Density)	0.342		0.742*		1.202**			
	(0.514)		(0.420)		(0.562)			
Ln(Weighted_Density)		-0.004		0.143**		0.248**		
		(0.109)		(0.072)		(0.120)		
Prp_Band13			0.036**	0.038**			0.026*	
			(0.015)	(0.017)			(0.015)	
Prp_Band6					-0.017***	-0.016***		-0.010**
					(0.006)	(0.006)		(0.004)
constant	-3.233*	-2.014***	-7.401***	-5.515***	-8.128***	-4.936***	-3.304***	-2.735***
	(1.832)	(0.600)	(1.527)	(0.876)	(1.596)	(0.729)	(0.794)	(0.672)
R2_Overall	0.912	0.909	0.932	0.916	0.916	0.883	0.917	0.903
Wald_Chi2	350.943	325.415	356.894	278.162	283.342	263.688	322.931	349.540
RESET_P_value	0.35	0.51	0.08	0.03	0.32	0.06	0.54	0.52
BPagan_Test_P_value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Serial_Corr_P_Value	0.00	0.01	0.00	0.01	0.02	0.05	0.00	0.02
StdDev_Ind_Effect	0.12	0.14	0.12	0.13	0.12	0.14	0.13	0.14
StdDev_Idiosy_error	0.08	0.08	0.08	0.08	0.08	0.07	0.08	0.08
Corr_comp_error	0.71	0.77	0.67	0.75	0.72	0.79	0.73	0.76
N_Sample_Size	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
T_Sample_Size	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00
Observations	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.00

Source: Economic Regulation, Thames Water.
* p<0.10, ** p<0.05, *** p<0.01

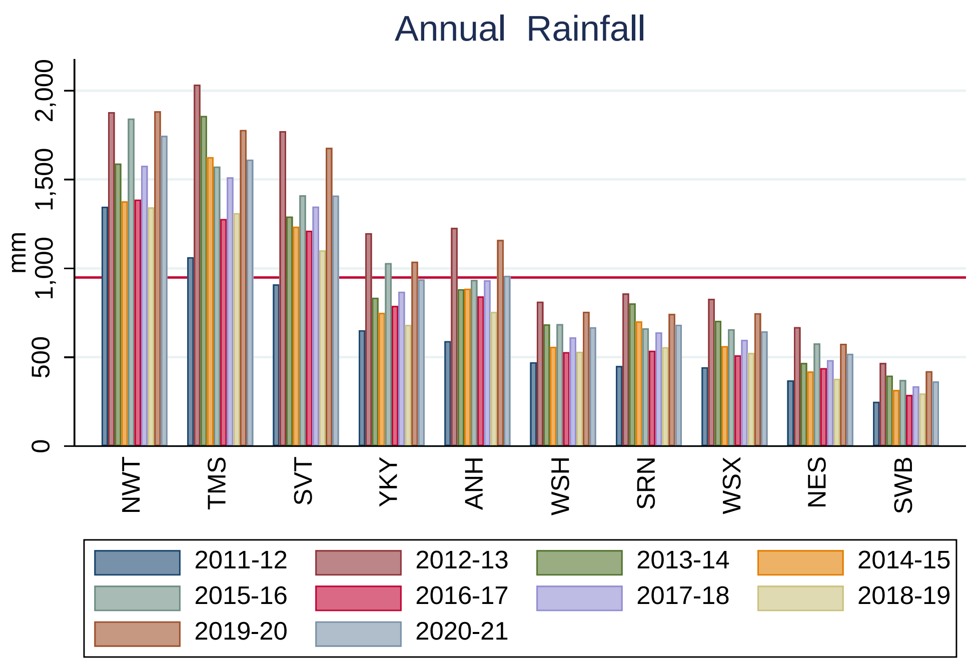
After this analysis we conclude that *band and density* drivers within the same model specification tend to be less robust in its R^2 and functional form specification when compared to the specifications that control for *band and complexity* at the same time (see models *Compl_Band*). The models with *complexity and density* yield some intuitive problems and the effect of density is not statistically significant due to the potential high degree of correlation between these two drivers as shown in the complexity and density Pearson correlation figure above.

Based on this evidence, we would like to suggest building a model for wastewater network that contains the ***load*** factor, the ***band size*** and the ***complexity dimension*** of the treatment. If density is included in the models, we think that this driver has to be controlled in a separate model with density and bands only, but not density and complexity, as the latter is quite unstable. Hence, a weighted process among these models could be explored. In other words, a good starting point to develop models for waste network plus are models WWWNP1 and WWWNP2, presented in Table 5 "*Potential Botex + PR24 Waste Network Models".*

New Potential Drivers to Explore

Within this context we have also been trying to incorporate what we think should be considered as drivers in the models for **environmental/climate change** we mentioned in the answer to question Q1. We think that a good candidate for this is the *annual rainfall* or a proxy that is related to a climate change impact[26]. In general, rainfall can be considered as an exogenous and out of companies' management control. Figure 3 below is a representation of the different levels of annual rainfall across the industry. ***We take this driver as an example only***, and not suggesting that this is the best way to approach it but considering this current information the models seem to show some significant impacts.

Figure 3 – Annual Rainfall



Source: Economic Regulation, Thames Water. Data Source: United Utilities

Before adding this driver into the network plus models, we want to add the annual rainfall into the sewage collection models to understand its impact at this part of the value chain. Table 8 below shows the sewage collection models with the annual rainfall using the CMA framework:

[26] We would like to thank United Utilities for their sharing in the data that contains annual rainfall and annual urban rainfall. We have considered both variables in the models, but we think that annual rainfall would capture also the infiltration issues faced in rural areas, and not just considering the urban element of the driver. We also think that the concept of rainfall incorporating a rural and urban factor and re-calculated by Ofwat should be an option to explore the impact that this driver can have on base costs.

Table 8 – Sewage Collection Models CMA with Annual Rainfall

Sewage Collection Models CMA with Rainfall (Sample: 2011-12 to 2019-20)

	SWC1 b/se	SWC2 b/se	SWC1_R b/se	SWC2_R b/se	SWC1_UR b/se	SWC2_UR b/se
Ln(Sewer_Length)	0.839***	0.830***	0.883***	0.856***	0.752***	0.720***
	(0.050)	(0.094)	(0.066)	(0.103)	(0.059)	(0.095)
Ln(Pumping_Cty)	0.291*	0.501**	0.343*	0.527**	0.337*	0.507**
	(0.170)	(0.200)	(0.179)	(0.224)	(0.184)	(0.203)
Ln(Property_Density)	0.976***		0.914**		0.871**	
	(0.374)		(0.424)		(0.444)	
Ln(Weigted_Density)		-2.683**		-2.803**		-2.866**
		(1.309)		(1.412)		(1.291)
(Ln(W_Density))^2		0.194**		0.202**		0.205**
		(0.086)		(0.094)		(0.085)
Ln(Annual_Rainfall)			0.132***	0.132***		
			(0.029)	(0.025)		
Ln(Urban_Rainfall)					0.130***	0.136***
					(0.026)	(0.024)
constant	-8.030***	4.845	-9.201***	4.075	-7.586***	5.825
	(1.332)	(4.638)	(1.524)	(4.951)	(1.577)	(4.658)
R2_Overall	0.934	0.913	0.938	0.918	0.936	0.922
Wald_Chi2	796.180	614.311	567.777	2986.557	555.188	2896.212
RESET_P_value	0.10	0.16	0.11	0.14	0.10	0.05
BPagan_Test_P_value	0.00	0.00	0.00	0.00	0.00	0.00
Serial_Corr_P_Value	0.01	0.01	0.06	0.07	0.06	0.07
StdDev_Ind_Effect	0.10	0.15	0.11	0.16	0.11	0.15
StdDev_Idiosy_error	0.09	0.09	0.09	0.08	0.09	0.08
Corr_comp_error	0.54	0.74	0.59	0.79	0.61	0.75
N_Sample_Size	10.00	10.00	10.00	10.00	10.00	10.00
T_Sample_Size	9.00	9.00	9.00	9.00	9.00	9.00
Observations	90.00	90.00	90.00	90.00	90.00	90.00

Source: Economic Regulation, Thames Water.
* p<0.10, ** p<0.05, *** p<0.01

Table 8 "*Sewage Collection Models CMA with Annual Rainfall*" seems to reflect the significant impact that annual rainfall has on the models. The statistical properties of the models hold and are stable with some marginal improvements. We have also included in the last columns of this table the impact of annual urban rainfall. We need to recognise that the driver was not quite robust when changing the sample periods. By changing the sample periods, we mean moving the sample from the initial years 2011-12 onward to 2012-13 and so on, but when changing the sample backwards, for example using the period 2011-12 to 2018-19, or 2017-18, the results were quite robust and that there is a high degree of correlation between Load and Rainfall, around 0.90[27]. These facts might be driven by some multicollinearity issues causing the instability of the driver across time or that the effect of rainfall is becoming more substantial the more the time progress due to climate change pressures. Table 9 below contains the models results for WWWNP and the extensions described before with annual rainfall.

[27] Although the Sewage Collection Model used by the CMA that includes Property Density is also not robust when changing the time sample period. This also applies to the PR19 Ofwat models. Hence, it would be very useful if Ofwat clarifies what is robustness of time periods, or what is expected in this checking part of the modelling.

Table 9 – WWWNP Extensions with Annual Rainfall

Starting Point for Waste Network Plus Econometric Models with Rainfall (Sample: 2011-12 to 2019-20)

	WWWNP1 b/se	WWWNP2 b/se	WWWNP11 b/se	WWWNP12 b/se	WWWNP13 b/se	WWWNP14 b/se
Ln(Load)	0.684***	0.708***	0.679***	0.722***	0.613***	0.653***
	(0.058)	(0.060)	(0.064)	(0.070)	(0.057)	(0.070)
Prp_NH3_Below3	0.004***	0.004***	0.004***	0.004***	0.004***	0.004***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
Prp_Band13	0.026*		0.023		0.023	
	(0.015)		(0.015)		(0.015)	
Prp_Band6		-0.010**		-0.011**		-0.010**
		(0.004)		(0.005)		(0.004)
Ln(Annual_Rainfall)			0.066***	0.069***		
			(0.022)	(0.022)		
Ln(Urban_Rainfall)					0.076***	0.075***
					(0.023)	(0.022)
constant	-3.304***	-2.735***	-3.689***	-3.312***	-2.907***	-2.499***
	(0.794)	(0.672)	(0.961)	(0.811)	(0.821)	(0.771)
R2_Overall	0.917	0.903	0.917	0.901	0.917	0.902
Wald_Chi2	322.931	349.540	311.278	319.201	327.636	323.937
RESET_P_value	0.54	0.52	0.51	0.41	0.50	0.40
BPagan_Test_P_value	0.00	0.00	0.00	0.00	0.00	0.00
Serial_Corr_P_Value	0.00	0.02	0.03	0.08	0.03	0.08
StdDev_Ind_Effect	0.13	0.14	0.15	0.15	0.15	0.16
StdDev_Idiosy_error	0.08	0.08	0.08	0.08	0.08	0.08
Corr_comp_error	0.73	0.76	0.77	0.80	0.77	0.80
N_Sample_Size	10.00	10.00	10.00	10.00	10.00	10.00
T_Sample_Size	9.00	9.00	9.00	9.00	9.00	9.00
Observations	90.00	90.00	90.00	90.00	90.00	90.00

Source: Economic Regulation, Thames Water.
* p<0.10, ** p<0.05, *** p<0.01

Finally, but not least, we have been also trying to extend the scope of consents to a measure that not only captures NH3 but also BOD and P. We have explored a potential weighted consent variable using the tightest consents for the three different determinants we have in the industry i.e., ammonia (NH3,<1mlg), phosphorous (P,<0.5mlg) and BOD <7mlg. Using the tightest consents could bring an objective approach on the selection of the thresholds of the consents for each determinant. We have also considered the industry weights for each of these determinants as the total load treated under each of these determinants as a proportion of the total load received of the industry. This weight would capture the intensity of the consent imposed by the Environmental Agency and its interaction with the companies' own specific circumstances in treating any determinant. The mathematical expression (1) below tries to reflect this[28]:

$$\begin{aligned} Composite\_Consent_{it} &= W_{NH3<1mlg,Industy} * \left(PrpNH3_{<1mlg,it}\right) + W_{P<0,5mlg,Industy} * \left(PrpP_{<0.5mlg,it}\right) + W_{BOD<7mlg,Industy} \\ &* \left(PrpBOD_{7mlg,it}\right) \qquad \mathbf{(1)} \end{aligned}$$

Where:

$$Industry\ Weigth = W_{NH3<1mlg,Industy,t} = \frac{\sum_{i=1}^{N} NH3_{<1mlg,it}}{Total\ Load_{Industry,t}} \qquad \mathbf{(2)}$$

[28] *Please note that this is just a proposal*. We believe that this measure can be substantially improved to get more accurate numbers.

Similar weights for P and BOD as previously shown in expression (2) for the case of NH3 industry weight consent. The company specific components are illustrated in expressions (3,4,5) below, where *i*=company and *t*=year:

$$Proportion\ Company = Ammonia = PrpNH3_{<1mlg,it} = \frac{NH3_{<1mlg,it}}{Total\ Load_{it}} \quad (3)$$

$$Proportion\ Company = Phosporus = PrpP_{<0.5mlg,it} = \frac{P_{<0,5mlg,it}}{Total\ Load_{it}} \quad (4)$$

$$Proportion\ Company = Biological\ Oxygen\ Demand = PrpBOD_{<7mlg,it} = \frac{BOD_{<7mlg,it}}{Total\ Load_{it}} \quad (5)$$

After doing the calculation of the previous expressions we named the variable "*Composite Consent*". We take the natural logarithm of this variable before it is introduced in the econometric models. Table 10 below label this variable as "*Ln(CompositeNH3PBOD)*". The econometric results are:

Table 10 – WWWNP Extensions with Annual Rainfall and Composite Consent

Starting Point for Waste Network Plus Econometric Models with Rainfall and Composite Consent (Sample: 2011-12 to 2019-20)

	WWWNP1 b/se	WWWNP2 b/se	WWWNP15 b/se	WWWNP16 b/se	WWWNP17 b/se	WWWNP18 b/se	WWWNP19 b/se	WWWNP20 b/se
Ln(Load)	0.684***	0.708***	0.744***	0.768***	0.679***	0.722***	0.745***	0.786***
	(0.058)	(0.060)	(0.068)	(0.085)	(0.064)	(0.070)	(0.071)	(0.096)
Prp_NH3_Below3	0.004***	0.004***			0.004***	0.004***		
	(0.000)	(0.001)			(0.000)	(0.001)		
Prp_Band13	0.026*		0.030**		0.023		0.029**	
	(0.015)		(0.014)		(0.015)		(0.015)	
Prp_Band6		-0.010**		-0.011**		-0.011**		-0.013**
		(0.004)		(0.005)		(0.005)		(0.005)
Ln(CompositeNH3PBOD)			0.042**	0.044***			0.042**	0.044***
			(0.017)	(0.015)			(0.017)	(0.016)
Ln(Annual_Rainfall)					0.066***	0.069***	0.069***	0.072***
					(0.022)	(0.022)	(0.022)	(0.021)
constant	-3.304***	-2.735***	-3.751***	-3.047***	-3.689***	-3.312***	-4.232***	-3.658***
	(0.794)	(0.672)	(0.968)	(0.936)	(0.961)	(0.811)	(1.127)	(1.124)
R2_Overall	0.917	0.903	0.934	0.911	0.917	0.901	0.935	0.907
Wald_Chi2	322.931	349.540	480.873	241.864	311.278	319.201	469.443	219.465
RESET_P_value	0.54	0.52	0.05	0.13	0.51	0.41	0.15	0.32
BPagan_Test_P_value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Serial_Corr_P_Value	0.00	0.02	0.00	0.01	0.03	0.08	0.02	0.04
StdDev_Ind_Effect	0.13	0.14	0.12	0.13	0.15	0.15	0.13	0.14
StdDev_Idiosy_error	0.08	0.08	0.08	0.07	0.08	0.08	0.08	0.07
Corr_comp_error	0.73	0.76	0.69	0.76	0.77	0.80	0.74	0.79
N_Sample_Size	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
T_Sample_Size	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00
Observations	90.00	90.00	90.00	90.00	90.00	90.00	90.00	90.00

Source: Economic Regulation, Thames Water.
* p<0.10, ** p<0.05, *** p<0.01

The wastewater network models WWWNP15 and 16 include the composite tight consent driver. There is an increase in the R^2 when compared to the models that use the proportion of ammonia NH3 below 3mlg (see models WWWNP1 and 2). The statistical test and level of significance suggest, overall a good performance of the models. Moreover, these models 15 and 16 seem to be also robust to some changes in the sample period, although not in all the specifications[29]. The last two columns of the table are an attempt to include jointly the effect of rainfall and the composite consent driver (WWWNP19-20). Overall, as it can be seen from some statistical tests in the table, the models WWWNP17 and WWNP18 provide some information on how much these models potentially could be extended. We do not want to suggest a model in particular with the inclusion of these two new potential drivers, but we want to highlight the potential areas of drivers where the models could be extended following the principles proposed in the consultation.

[29] We can provide the results if required.

Moreover, we also believe that these two new drivers can be substantially be improved in the way they have been measured in this proposal.

To sum up, our recommendation is to start developing a model from specifications WWWNP1 and WWWNP2, with the potential to include one more driver related to the environmental/green dimension proposed in this response (see response to Q1) and to carefully consider a potential composite consent driver. If the latter composite variable is not robust enough, we would suggest keeping the current proportion of NH3 below 3mlg in a potential WWWNP model, as this driver is also robust based on the initial analysis presented in this response. Lastly, the efficiency scores for these two initial WWWNP models versus the CMA and PR19 Ofwat's ranks are shown in Table 11 below:

Table 11 – WWWNP Efficiency Ranks

	Network plus Ofwat PR19 (Separate SWC+SWT)		Network Plus CMA (Separate SWC+SWT)		Network Plus using Econometric Models (WWWNP1-2)	
	Company	Efficiency score	Company	Efficiency score	Company	Efficiency score
	SVT	0.87	SVT	0.87	SVT	0.79
	WSX	0.87	WSX	0.91	WSX	0.85
	YKY	0.96	TMS	0.96	SWB	0.91
	ANH	1.00	SWB	0.99	TMS	0.96
	SWB	1.00	YKY	1.01	WSH	0.96
	TMS	1.00	ANH	1.02	ANH	0.96
	NES	1.03	WSH	1.03	YKY	0.97
	WSH	1.07	NES	1.04	NES	1.01
	SRN	1.07	SRN	1.17	NWT	1.15
	NWT	1.30	NWT	1.26	SRN	1.31
Mean		**1.02**		**1.03**		**0.99**
UQ		**0.97**		**0.97**		**0.92**

Source: Economic Regulation, Thames Water

The efficiency scores positions across the industry remains the same for the most efficient company and the second best. Below that, there are different changes based on the different model specifications and time periods that these results are based on. There is also a significant drop on the Upper Quartile (UQ) frontier when the model is assessed trough an econometric model such as the ones proposed in WWWNP1 and WWWNP2[30]. This last result on the UQ might suggest that there is still some scope to include other drivers such as the ones proposed in this answer, rainfall, or a composite consent.

Cost drivers and explanatory variables

Q9: Do you think we should reconsider the inclusion of APH in the wholesale water base cost models at PR24? If so, should it be a substitute for, or additional to, booster pumping stations per length of mains?

We support the inclusion of APH in the wholesale water models. There is substantial engineering support to use this driver in the water models and it has been used in several price reviews in the past. More details on our support to this driver can be found in our response to the "*PR24 and Beyond*" consultation in May 2021, question Q10.1. In addition, this variable has more characteristics of an exogenous driver i.e., potentially free from management control.

[30] We weighted these two models with 50%.

We also consider that APH should be used instead of booster pumping stations per length of mains and the latter should be used only when all possibilities to include APH have been explored.

Q10: Should we consider replacing the existing 'load treated in size band 6' variable with 'load treated in band 8 and above' in the relevant wholesale wastewater base cost models?

Changing the bands could bring some practical problems.

One problem could be related to the creation of unnecessary outliers, which would only add more issues in models that are already restricted for other technical issues such as the sample size in waste.

A second problem would be the definition of the thresholds to define different bands, which is a similar problem as the one faced with the Density of Dispersion variable that was attempted at PR19. Hence, the threshold and some subjectivity can be added in the construction of the variable.

Q11: Please provide detailed proposals for any additional / alternative cost drivers and explanatory variables we should consider at PR24, including clearly defined data requirements that would need to be collected from companies.

Climate Change/Environmental Drivers

- Rainfall Rates: This could be seen as the one proposed by UU at PR19 where the annual urban runoff was explored. This UU proposal could be complemented by countryside rain component that may result in increased infiltration. This driver is exogenous and out of companies' management control. The data and information required to construct it could be based on the sources of external data used by UU with the complement of countryside rain.
- Winter/Summer Temperatures: Extreme weather conditions can have notorious impacts of costs in specific parts of the year. We consider that winter temperatures put more pressure to water companies than temperatures in other parts of the year. This exogenous driver sometimes is seen in unprecedented records, such as the beast from the east event in 2018. These extreme temperatures are quite intense in different areas and at different rates across years. These types of events are not easily controlled by companies or planned in advance. Moreover, summer temperatures have an impact on algal blooms and then on treated water processes. We think that using the regional information of the MetOffice could allow us to construct a variable of such environmental/climate change dimension.

Wastewater

- Total No of Civil Assets (Concrete and GRP) (Sewage Treatment).
- Total No of Mechanical assets (Sewage Collection): Expenditure related to these assets was historically not included in the botex, as they have a long life. However, we expect that these assets will start needing maintenance in the near future. Therefore, they are likely to incur botex, which should be reflected by the number of such assets. It may be that the scale variable captures some of this effect yet

including the number of these assets may be appropriate as size of STW and number of assets is not necessarily appropriate. Companies should have a good idea of how many assets of each of these types they have held historically.

- **Proportion of sewage which flows through combined sewers (rain/sewage mix):** This proportion would influence the cost, as the sewage which flows through combined sewers would lead to greater sewage collection pumping costs per population served, as well as a greater utilisation of storm tanks. This can also increase the likelihood of storm discharges. This will be increasingly important if we are looking at partially treating all sewage. Ofwat does not currently collect this data, however it may be something to be considered for future modelling, especially as we move forwards with the industry goal of reducing sewage spills.

Model Estimation Method

Q12: Do you agree that we should maintain the use of random effects to estimate our wholesale base cost models at PR24?

The outcomes of the econometric models in water and waste have shown the presence of an unobserved heterogeneity consistently and robustly across all models estimated at PR19. The models we have explored in question Q8 regarding the development of a wastewater network model and some other examples in treated water distribution explored in question Q22 also go in the same direction of the random effect (RE) models. This was also the case in the CMA models. This has been confirmed by the Breach-Pagan test for RE[31] that helps to understand the presence of this unobserved heterogeneity when compared to a pooled model error term structure. In conclusion, there is a systematic presence of an unobserved heterogeneity across all botex models.

The next question that can be asked is the choice between Fixed effects (FE) and Random effects (RE) models. Some researchers tend to use the Hausman test rejection as an indicative form to reject the FE models and a non-rejection as the adoption of RE model. Some other authors prefer to think the issue between FE and RE as an issue of model selection rather than hypothesis testing[32]. The key question to assume a RE model is to believe that none of the cost drivers in the models are correlated with the unobserved heterogeneity or to assume exogeneity of all the regressors with the unobserved time-invariant individual effect. For example, if a service driver is included in the models, we would need to understand to what degree is this driver correlated with the error term. In most of the models used so far to assess botex costs, the main assumption on cost drivers sits on the exogeneity of all drivers in the short-run. If that is the case and given the low degree of within variation among most of the cost drivers used in the models, the RE models and its Feasible GLS estimator (FGLS) seems to be the most appropriate model and estimator to use in the estimation of the models. We also welcome the clarification of the type of RE model we are considering for the botex models if it is a one-way error term model or a two-way RE error term model[33].

Lastly, as mentioned in our answer to question Q5, it is important to recognise the sample size and its potential effects on some of the different components needed to implement the estimation

[31] The null hypothesis is $Ho: \sigma_{\alpha}^{2} = 0$ which basically test for the variance of the unobserved heterogeneity to be zero.

[32] See Hsiao and Sun (2000). To pool or not to pool panel data. Chapter 9 in, Panel Data Econometrics: Future Directions, *North-Holland*.

[33] By the two-way RE model, we mean the specification of the error term as $u_{it} = \alpha_i + \lambda_t + e_{it}$ where α_i is the time invariant unobserved heterogeneity, λ_t unobserved time effect such as the impact of regulators laws etc. and e_{it} is the random error. A one-way error term, which is what we understand we are currently using in the botex models is the same specification of the error term u_{it} above but without the unobserved time effect.

of RE models. For instance, in small samples and unbalanced panels[34] it is useful to consider or contrast the outcomes of the models with the small-sample Swamy–Arora estimator individual-level variance component that can be used instead of the default consistent estimator. Swamy–Arora estimator of the variance includes a degree-of-freedom correction to improve its performance in small samples. This can bring some improvements in efficiency and performance in small samples[35].

Within this context we believe that the RE model is the appropriate model for base costs, based on the evidence provided by statistical tests, characteristics of the data and model assumptions.

Model Selection Process

Q13: Do you agree with our proposed model selection process?

We agree with the proposed selection process, the tests and in general the sensible steps, such as the sign and statistical level of significance etc. However, as is mentioned in the consultation not all tests would be achieved in a model or are entirely clear. We believe that it is important to be flexible on some of the tests to be used in the assessment of models. For instance, a certain degree of multicollinearity could be admitted if other benefits are gained in the model, or a more marginal level of statistical significance can also be admitted in a particular model, if overall the model performs well. However, some tests are more critical than others and the test should be applied with more rigour where this is the case, for example, misspecification of the functional form. This could be more critical than the previous tests or assessment requirements, particularly when assessing efficiencies. A missing driver or misspecification of the functional form could have a major effect on efficiency scores.

In addition, we would encourage Ofwat to set out ranges or provide more clarity on what is admissible for a model to be considered as “good”. For example, when assessing a model under different sample periods we consider that it should be clear what exactly this means. In the PR19 and CMA models, some models, for example, in sewage collection were not robust enough when changing the sample period, although they were used to determine allowances. Table 12 below shows this situation for one of the sewage collection models where it is appreciated that the effect of pumping capacity per length of mains disappears when the sample moves from the initial years 2011-12 and onwards[36]:

[34] Clearly, water is an unbalanced panel and waste could also be the same if we have, for PR24, around 11 companies.

[35] This could be easily implemented in Stata by the command option **sa** that represents the Swamy-Arora estimator, particularly in unbalanced panels like water and potentially waste.

[36] These results are based using the data set used by the CMA and moving the beginning of the same onwards (2012>) as well as moving from the end of the sample backwards (<2020).

Table 12 – Sewage Collection with different samples

Sewage Collection Models CMA with Different Sample Periods

	SWC1_CMA b/se	SWC1(2012>) b/se	SWC1(2013>) b/se	SWC1(2014>) b/se	SWC1(2020<) b/se	SWC1(2019<) b/se
Ln(Sewer_Length)	0.839***	0.840***	0.844***	0.863***	0.845***	0.822***
	(0.050)	(0.059)	(0.053)	(0.053)	(0.055)	(0.042)
Ln(Pumping_Cty)	0.291*	0.261	0.254	0.268	0.326*	0.290**
	(0.170)	(0.179)	(0.172)	(0.181)	(0.182)	(0.130)
Ln(Property_Density)	0.976***	0.847**	0.918**	0.947**	0.980***	1.129***
	(0.374)	(0.432)	(0.404)	(0.381)	(0.355)	(0.201)
constant	-8.030***	-7.540***	-7.850***	-8.180***	-8.133***	-8.417***
	(1.332)	(1.535)	(1.430)	(1.312)	(1.334)	(0.956)
R2_Overall	0.934	0.935	0.938	0.939	0.931	0.941
Wald_Chi2	796.180	535.664	604.613	527.406	680.794	654.208
RESET_P_value	0.10	0.09	0.09	0.09	0.10	0.10
BPagan_Test_P_value	0.00	0.00	0.00	0.00	0.00	0.00
Serial_Corr_P_Value	0.01	0.01	0.01	0.01	0.01	0.01
StdDev_Ind_Effect	0.10	0.10	0.11	0.11	0.10	0.10
StdDev_Idiosy_error	0.09	0.08	0.08	0.08	0.09	0.08
Corr_comp_error	0.54	0.60	0.61	0.67	0.57	0.60
N_Sample_Size	10.00	10.00	10.00	10.00	10.00	10.00
T_Sample_Size	9.00	8.00	7.00	6.00	8.00	7.00
Observations	90.00	80.00	70.00	60.00	80.00	70.00

Source: Economic Regulation, Thames Water.
* p<0.10, ** p<0.05, *** p<0.01

Moreover, we would also welcome clarification from Ofwat on what degree of acceptance on the RESET test is admissible, 5% or 10%, when assessing for functional forms. We think that both should be considered as valid practice.

4. Cost Adjustment Claims

Q14: Do you agree that the cost adjustment claim process at PR24 should be separated between base (wholesale and residential retail) and enhancement claims?

We believe this is a sensible approach and clarifies the basis for adjustment of costs. This effectively happened in PR19 where Ofwat assessed our metering cost adjustment claim as part of the metering enhancement case. We welcome Ofwat's decision to retain this distinction.

We recognise CACs are an important part of any cost assessment process for both base and enhancement expenditure, and we would like Ofwat to note that setting the hurdle bar too high is not in customers' interest. If companies are to invest to improve services for customers, communities and the environment they need to be properly funded, including for costs that are not properly covered in the models and instead covered by cost adjustment claims.

At PR19 the materiality bar for Thames Water was around £40m per claim. We suggest that this limit should be applied at the price control level rather than at a claim level as several claims short of the materiality level could accumulate to a substantial amount. This could be associated with a triviality threshold per claim to avoid immaterial issues.

Q15: What base cost adjustment claims (wholesale and residential retail) would you consider submitting if the PR19 base cost models were used to assess efficient costs at PR24?

Our business plan for PR24 is at an early stage of development and our understanding of the required investments is expected to change over the next months. At the moment, we believe cost adjustments claims (CAC) may be required in a number of the areas of our wholesale water and wastewater plan:

- Replacement of distribution mains.
- Maintenance costs of our trunk mains network.
- Operational Technology (OT).
- Maintenance of large reservoir assets (e.g., Queen Mother)
- Renewal of long-life assets never replaced before (e.g., civils)
- Maintenance and safe operation of our wastewater network.
- Regional wages.

In addition, we will have costs associated with operation of the Thames Tideway Tunnel, which has traditionally been included within a separate price control with a separate bespoke cost assessment. We assume that Ofwat will again use a separate price control, although this will need a licence change, and we would welcome a discussion with you as to how you wish to assess these costs at PR24.

Q16: What additional cross-sector data should be collected to support the submission of the claims indicated in response to the previous question? Please describe and explain the rationale behind the additional data that you consider should be collected and provide a draft definition

We set out below some potential areas for additional data collection. We would be happy to work with Ofwat and the industry to develop these into more detailed specifications as required.

Trunk mains are a source of significant and we consider exceptional maintenance cost for Thames Water. The collection of this data would show whether this was true or not and enable calculation of a symmetrical adjustment, if required.

- Lengths of trunk mains by size – 18”; 24”; 36”; 42+”.
- Maintenance costs per km.
- Number of customers connected per km.

We are likely to have significant increases in operational technology expenditure due to age replacement and changes resulting from BT and Government decisions. It is not clear to what extent companies have been incurring material costs to replace these types of assets in recent years. Collecting data on replacement expenditure on these assets would demonstrate the extent which these costs are included in the botex modelling or whether some adjustment is required.

Costs associated with maintenance of large reservoir assets maybe adversely affecting the ability to create reliable Water Resources models. We therefore suggest that data on number and type of large reservoir assets and the costs associated with maintenance of these assets would be useful to either improve water resources or enable a symmetrical adjustment to be estimated.

As highlighted in our response to the May 2021 consultation PR24 and beyond, there are issues with the data underlying the analysis of regional wages. Better data on the Standard Occupational Categories (SOC) from companies would enable more conclusive evidence as to whether there needs to be an adjustment for regional wages and potentially a symmetrical adjustment.

Q17: How can the cost adjustment claim guidance be enhanced to improve the quality of cost adjustment claim submissions?

We appreciated the Ofwat engagement on the topic of Cost Adjustment Claims, through the CAWG, which we see as an important part of the cost assessment process. We believe that early publication of botex models will help improve the quality of our CACs for PR24. We also appreciate Ofwat's effort to provide greater clarity on the expected evidential threshold as part of the recent Cost Assessment Working Group (CAWG) on CACs. The guidance provided is helpful and we believe this should enable water companies to improve the quality of CACs for PR24.

However, we reiterate that it is important that the evidential and materiality levels are not set too high. The use of symmetrical adjustments for cost adjustment claims also needs to be carefully introduced with recognition that in some cases it will not be appropriate and/or practical. If companies are to invest to improve services for customers, communities and the environment they need to be properly funded to do so.

Q18: Would an early cost adjustment claim submission be welcome at PR24?

We believe early CACs submission is possible in principle, although it should not replace a final submission as part of the formal business plan submission. However, when considering this opportunity, we believe Ofwat should consider the issues that come with it. A company ability to submit a CAC can depend on its individual circumstances, e.g., involvement in a bespoke regulatory process like the LWI conditional allowance, which might not be sufficiently mature by the time Ofwat requires early submission. Similarly, where there's a new evidential requirement to be met, e.g., in the case of symmetrical cost adjustments, Ofwat should take into account the additional activities (e.g., data requirement) that companies are required to carry out in order to provide the necessary evidence. Ofwat should allow sufficient time for companies to do this, to avoid restricting a company's ability to submit a valid claim. We believe issues need to be properly accounted for in the submission process. Whether or not there is an early submission required, we believe it would be useful to have an early meeting to discuss potential CACs ahead of submission.

5. Capital Maintenance and Asset Health

Q19: Do you agree with the different elements / approaches to introducing more of a 'forward-look' into our approach to assessing capital maintenance expenditure? Are there other elements / approaches we could consider?

We have commented on the approach of including Include forecast costs in the wholesale base cost econometric models and setting a forward-looking catch-up efficiency challenge in our response to Q5 and Q6.

Another method would be to consider amending cost drivers although this may be difficult in the short term. While we welcome Ofwat's consideration of asset health measures in this consultation, there is a need for more significant changes to increase levels of funding for capital maintenance. Current levels of asset renewal with replacement rates of 0.2% and 0.6% for sewers and water mains respectively are not sustainable and we consider that more urgent action is needed, with increased allowances for capital maintenance asset renewal in PR24.

The current incentives strongly discourage companies from including more botex expenditure in their business plans than they expect Ofwat's models to allow. So even if Ofwat consider including forward projections in their modelling and consider asset health indicators there will be significant barriers for companies to include what they really consider is needed in their business plan submissions.

However, it is essential that we start the process of increasing capital maintenance and moving replacement rates to more sustainable levels in AMP8. We would encourage Ofwat to ask companies to make an *"enhancement case", for additional capital maintenance* in their business plan submissions, that will not have any negative consequences on cost sharing rates and other business plan incentives if ultimately rejected by Ofwat. Ofwat could consider these plans alongside the forward-looking asset health metrics and econometric modelling, to assess efficiency and provide a separate additional allowance.

To ensure companies did not gain from this specific allowance it could be linked to clear activity levels, with money returned to customers if the essential activity was not undertaken. There would also need to be an adjustment to botex allowances, to reflect the implicit allowance and avoid double counting, and to PC levels as appropriate to provide customer protection. Companies would have to justify the appropriate levels of capital maintenance to reflect their local circumstances.

At the very least we consider Ofwat should trial this approach in PR24 with mains replacement, which is widely recognised as an area that all companies should be increasing.

Q20: Do you have any comments on the proposed long list of asset health measures in Table 5, particularly in relation to their suitability and how feasible they are to collect? Please include any reporting or definition changes you would like us to consider and provide suggestions for other measures not included in this list.

We set out below our comments on the longlist of asset health measures.

1. Summary

We support the introduction of a wider range of asset health measures to enable them to form a holistic and more complete view of the health of assets across the sector.

Ofwat's proposal to develop robust forward looking asset health metrics is aligned with our Asset Health Policy which aims to ensure we understand the health of our assets and how to manage them thereby reducing the risk of current and future asset and service failure. Our overarching approach is to develop a suite of leading and lagging metrics which inform asset health across all key components of the water and wastewater asset value chain. We will use the asset health metrics to track the overall health of the assets, identify early indicators of asset deterioration and identify assets that may pose a risk to current and future service performance.

We have assessed the initial list of potential additional asset health measures as well as the draft list of asset health measures developed by the UKWIR Future Asset Planning project. Considering our current asset health data collection capability and the level of difficulty to start collecting additional data, we are supportive of the use of the following:

- Unplanned maintenance which we consider to be better suited to non-infrastructure assets. Currently we only collect emergency defect work at location level.
- Asset condition grade accompanied by Asset inspections planned vs actual. It is important to note that we currently collect asset condition data for non-infrastructure

mechanical and electrical (M&E) assets and a defined group of critical asset cohorts, e.g., sewers in the rail environment. Outside of the critical asset inspection programmes we do not carry out regular inspections of our gravity sewers to monitor deterioration.

- Mean Time Between Failures is a good measure of asset health however we do not currently collect reactive maintenance at a level to enable reporting of this measure.
- Base asset health index (BAH) to enable a forward-looking view of the level of asset health at asset cohort level or regionally. We would need to establish systems and reporting and allow collection of a number of years data for this measure to be meaningful.

We would welcome further discussions with Ofwat regarding the reporting granularity of these measures. We would support the reporting of these measures at an asset cohort level to ensure that the detail around these cohorts is not lost at a higher level of aggregation.

Initially the focus of the Asset Health measures should be on asset condition, failure rates and maintenance activities as these will be easier to report, although the adoption of some of the measures will require amendments to our data collection and reporting activities.

Over time we would recommend that asset performance measures be introduced to complement the condition type measures to give a fuller picture of the health of the assets at a cohort level. At all times we need to be mindful that these measures should be both leading and lagging. Introducing such measures will require further discussion with both Ofwat and the industry and are likely to require major changes to both data collection and reporting.

The definitions and methodologies for the new measures will need to be prescriptive to ensure the reporting is consistent across all companies.

2. Asset Characteristics

1.1 Asset condition grade (for assets or groups of assets)

We consider this measure to be suitable for asset health reporting. As with all of these suggested measures it is important to have a clear definition and scope of the measure to ensure consistent reporting across companies and for us to comment on the ease of collection.

We collect condition grade data on our non-infrastructure mechanical and electrical assets when a technician carries out a maintenance task through our work order process. This is supported by condition-based monitoring on a small selection of our most critical assets.

Regular planned CCTV surveys are carried out on some of our critical gravity sewers through our sewers in the rail environment programme, outside of this programme we do not conduct regular CCTV surveys on the same sections of sewers to aid deterioration monitoring.

We also undertake visual condition assessments as part of our programme of inspections of tunnels, service reservoirs, storage reservoirs and the Thames Water Ring Main. We also have a CCTV programme to survey our boreholes.

The “Asset inspections planned vs actual” measure proposed under the maintenance activity type will complement this measure by demonstrating the rate that condition grade data is updated. It should be noted that the asset inspections measure on its own is not a suitable asset heath measure as it does not describe the condition of an asset.

3. Maintenance Activity

3.1 Unplanned maintenance

Unplanned maintenance is one of the key maintenance indicators of asset health. In order to make this measure successful it needs to be clearly defined as there are a variety of ways of measuring this type of activity.

Currently we only collect reactive work data for our non-infrastructure assets at location level.

Reactive work on our sewer network is primarily captured through the sewer collapse and blockage performance commitments.

For our water assets we capture visible (reported) mains repairs as part of the mains repairs performance commitment and sometimes there are linkages between unplanned maintenance and the unplanned outage performance commitment. We also capture visible leakage activity numbers on communication pipes and ancillaries.

3.2 Planned network rehab

For water infrastructure, planned network maintenance was one of our sub-measures for the water infrastructure asset health measure. We think there is merit in adopting a planned network rehab metric for sewers, rising mains and water mains at PR24, if the Botex models could be adapted to take the cost of this work into account.

Currently, the amount of planned network rehab delivered by a company can be influenced by many factors, such as a company's risk appetite, the age and criticality of the assets, the design of the networks and their location. This will lead to variances across companies with this type of measure and may lead to difficulties in making comparisons, however using this measure over the longer term will demonstrate how individual companies are managing their asset deterioration.

We would suggest normalising this data against the overall size of the network for this measure to become useful.

3.3 Proactive vs reactive maintenance

This is a useful measure as it allows you to understand the impact of the delivery of planned maintenance on assets and how companies are using maintenance activities to ensure an asset reaches or exceeds the required asset life. However, in some cases doing more planned maintenance does not always lead to a marked reduction in reactive work as assets can fail due to other reasons such as external factors (weather, customer behaviour) or the original design and installation of the asset leads to a higher failure rate.

For non-infrastructure assets, it is important for us to understand the granularity of data required for this measure as previously mentioned we only collect reactive work data for our non-infrastructure assets at location level.

3.4 Mean Time To Repair, Mean Time Between Failures

Mean Time Between Failure is a good indicator of asset health as it demonstrates failing health through a reduction in the time between failures or the effectiveness of planned maintenance in offsetting deterioration. Since we only collect reactive work data at location level, we are currently unable to report this measure. We are currently reviewing our data and systems to see what improvements are needed to enable the reporting of this measure. This measure is better suited to non-infrastructure mechanical and electrical assets.

Mean Time To Repair is a less suitable measure as return to service can be impacted by the level of asset redundancy deployed, the use of temporary plant or the availability of maintenance resource.

3.5 Maintenance backlog

We believe this measure is not a suitable indicator of asset health as it is a measure of the availability of maintenance resource, the work management priority applied to activities and the level of asset redundancy deployed at a location. The size of the backlog can also be influenced by weather and other external factors such as the speed of response from the supply chain.

3.6 Asset inspections planned vs actual

As mentioned in the asset characteristics section above, tracking the delivery of asset inspections complements the asset condition grade measure as it demonstrates that condition grade data is regularly being updated and deterioration tracked.

This measure like others would need to be clearly defined to ensure consistency of approach across the industry, we would recommend excluding items such as infrequent structural condition assessments and focus more on frequent regular condition grade monitoring.

4. Asset and Service Performance

4.1 Compliance Risk Index (CRI) (water treatment works, supply points, service reservoirs and water supply zones sub-measures)

We currently report on this measure.

4.2 Properties at risk of receiving low pressure

We currently report on this measure as an asset health metric and can continue to do so.

4.3 Sewer blockages

This measure does not adequately represent the deterioration and health of sewers. The root cause of the majority of blockages is customer behaviour and sewer misuse, not asset condition. We consider sewer collapses to be a more suitable measure of sewer asset health and performance given that it is driven by asset deterioration. However, this is a lagging indicator and we would recommend that it is complemented by the use of the Planned Network Rehab measure which will provide a more forward looking view of the impact of deterioration and companies plans to address this. We consider sewer defect density as proposed by the UKWIR Future Asset Planning project to also be a suitable leading indicator of sewer asset health. We only carry out regular CCTV inspections of a very small proportion of gravity sewers (e.g., sewers in the rail environment).

4.4 Percentage of population equivalent, served by sewage treatment works with numeric limits, which were non-compliant with: sanitary look-up table limits or nutrient limits, urban wastewater treatment directive (UWWTD) look-up table limits or nutrient limits

This measure is not an appropriate indicator of asset health at sewage treatment works as it focuses on the compliance with the discharge quality standards set out in a site's environmental permit. Discharge compliance can be affected by other factors such as human error, weather or illegal discharges into the sewer network. Poor asset health can also be masked by the use of additional manpower, tankering or the installation of temporary plant in order to reduce the risk of non-compliance caused by deterioration.

4.5 Number of equipment failures

We assume that this measure applies to failures of equipment on our networks such as pumping stations, penstocks and water booster stations. We consider this measure to be suitable for understanding the underlying trends in the failure of asset types and providing an indication of the rate of deterioration of asset health. We would need to understand which asset groups this measure would apply to be able to comment more fully.

4.6 Unplanned Interruptions greater than 12 hours

We think that the incentives to improve supply interruptions through asset health are already appropriately captured in the supply interruptions common performance commitment, such that this measure is not necessary.

4.7 Disinfection, Reservoir or Process control Index

We can report on this measure and it is supported by the UKWIR proposed optional measures.

5. Aggregated measures

5.1 Base asset health index (BAH)

We believe that this is a useful metric for assessing asset health at asset cohort and system level. It provides a systemic approach that would help inform future decisions on investment choices and risk using a forward-looking common view on remaining asset life.

Our aspiration is to develop BAH index methodologies for specific and prioritised asset cohorts by the end of AMP7, this will enable us to prioritise investment and manage risk in a more informed way. Gaps in our data to support the development of the BAH's will inform our asset information strategy and data quality improvements. A clear definition of this measure will be required to drive consistency across companies.

We believe that this metric could become a common asset health measure in the future, but not for AMP8. We need to establish systems and reporting and allow collection of a number of years data for this measure to be used in a meaningful way.

5.2 Overall Equipment Effectiveness (OEE)

We do not believe that this measure is suitable for asset health as it is a measure of asset efficiency. Whilst there are links between efficiency and asset health the efficiency of unhealthy asset(s) can be masked by the level of redundancy deployed or the use of temporary plant.

5.3 Asset risk (monetised likelihood*consequence)

This measure is not suitable as an indicator of asset health for the following reasons:

- It requires an industry standard methodology to monetise risk.
- Likelihood and consequence are subjective measures and risk scoring on these will vary due to bias factors coming into play, the level of redundancy deployed and the availability of suitable mitigation such as rezoning, tankering or temporary plant.

This measure will be difficult to apply consistently across the industry.

6. Additional measures

We are assuming that AMP7 common asset health performance commitments such as mains repairs, unplanned outage and sewer collapses will be rolled forward and included in the suite of asset health measures for AMP8.

Following a review of the UKWIR proposed asset health measures we believe that the following water measures should also be considered for inclusion:

- Trunk main wall thickness.
- Life expectancy estimated from wall thickness.
- Repairs of customer-reported defects.

6. Cost-Service Relationship

Q21: Do you agree with the high-level approach to determine 'what base buys'? Can you define any additional analysis or information that could support this process?

We welcome Ofwat's ambition in drawing a more explicit link between costs and outcomes for setting PR24 cost allowances and performance levels. Recognising the complexity of the cost-service relationship, we are supportive of taking a pragmatic and proportionate approach.

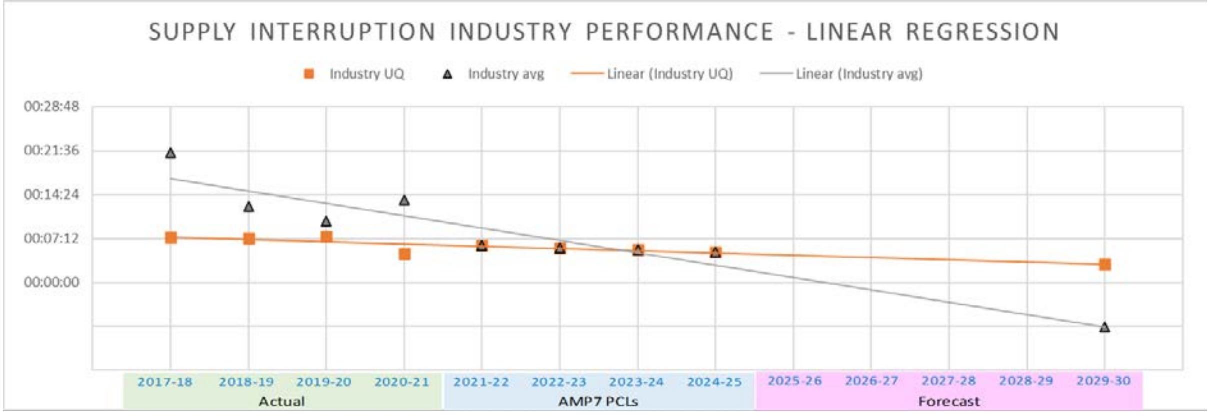
We do not consider the proposed high-level approach to determine 'what base buys' is *recognising the cost-service relationship*:

- The approach to forecasting PR24 baseline performance levels is completely separate from assessing the base cost.
- The assumptions underpinning the proposed approach do not reflect the reality that no company is delivering all upper quartile performances with upper quartile efficient cost.

We recommend forecasting performance levels from base for the upper quartile botex 'efficient' companies, to establish the central performance levels that botex 'efficient' companies would deliver. This can be compared with the industry upper quartile performance levels from base forecasted across all companies. The differences between the two should be considered at the package level rather than at individual performance commitment, recognising no company is delivering all upper quartile performances with upper quartile efficient base cost. The central performance levels from base forecasted on upper quartile botex 'efficient' companies should take precedent over the industry upper quartile performance levels from base forecasted across all companies when determining 'what base buys' for common performance level PCs.

Many AMP7 common PCs only have consistent performance data since 2017-18 shadow reporting. We are very cautious of applying linear regression forecast based on such limited data points. Together with the proposed assumption on using AMP7 PCLs as outturn performances, the linear regression approach could create absurd results. Figure 4 below is a linear regression example using supply interruptions actual performance during 2017-18 to 2020-21, then assuming AMP7 PCLs as outturn performances for the remain years in AMP7, as proposed by the consultation. The forecast industry average performance in AMP8 becomes much more stretching than the forecast industry upper quartile performance. This highlights the flaws of linear regression approach with small samples and the over simplistic assumption on future outturn performances.

Figure 4 - Supply interruption - linear regression



Source: Economic Regulation, Thames Water.

Whatever approach is used, we strongly encourage Ofwat to set out an adjustment mechanism in the PR24 final methodology that can take account of AMP7 actual delivery (and updated forecasts) when it becomes available at the later stage of PR24 and make adjustments to PR24 performance commitment levels if AMP7 actual delivery or updated forecasts are materially different from the assumptions applied in determining *"what base buys".*

The PR24 outcomes framework aims to streamline the performance commitments (PCs) package through increased focus on common PCs and far fewer bespoke PCs. Therefore, we consider the effort in establishing cost-service relationship should focus on eligible common PCs.

We provide further thoughts and suggestions on the key assumptions underpinning your proposed approach to better reflect actual delivery and cost-service link.

- Assumption 1 -'The performance level achieved by base should be set as a common performance level by default':
    - There are three AMP7 common PCs (supply interruptions, internal sewer flooding and pollution incidents) having common performance levels as targets across companies. The remaining AMP7 common PCs and comparable bespoke PCs have applied company-specific approaches in the target setting (including company-specific deadband). In 2020-21, half of the industry delivered targets for supply interruptions and pollution, less than a third of the industry delivered the internal flooding target.
    - We see the potential to expand 'common performance level' approach to other service and asset health metrics that have reasonably sufficient data to access the feasibility and effectiveness of such approach.
    - We recommend taking a top-down evaluation first to establish the feasibility of applying common performance level based on data availability and quality, performance drivers, and achievability.

- **Assumption 2** -'On average we consider that efficient companies will deliver their PR19 performance commitments':
    - This is not what we observed in 2020-21 actual performance for common PCs: 15 out of 17 companies had net ODI penalties from common PCs, and the sector had over £100m net penalty from common PCs[37] .
    - We acknowledge that when Ofwat publish PR24 final methodology and initial base cost and PCLs in December 2022, there will only be 2 years AMP7 actual delivery data available.
    - If using AMP7 PCLs as assumed outturn performances to forecast performance levels for base, (although we think making use of companies forecasts included in the APR reporting would be better) it is necessary to have a mechanism in the PR24 final methodology that can take account of AMP7 actual delivery (and updated forecasts) when it becomes available at the later stage of PR24 and make adjustments to PR24 performance commitment levels if AMP7 actual delivery or updated forecast is materially different from the assumptions applied in determining what base buys. This mechanism will also help to refine 2024-25 'year 0' baseline performance level, as Ofwat can obtain updated forecast on 2024-25 performances through companies' PR24 business plan submission in October 2023 and through 2023-24 annual performance reporting in July 2024.

- **Assumption 3** –'efficient companies will continue to improve performance over the long term from base expenditure':
    - In PR19, Ofwat suggested there was a 'positive correlation' between cost and service ((i.e., costs decrease as service quality improves), evidenced by a scatter plot chart mapping company performance against a basket of measures (quality rank) and a measure of cost efficiency (efficiency rank).
    - In the CMA's PR19 redetermination, it stated *"Ofwat said that its scatter plot analysis 'suggests a positive correlation'. We do not agree that this scatter plot provides robust evidence to support Ofwat's view. Simple statistical analysis shows that the correlation coefficient is not statistically significant from zero."*[38]
    - For leakage, the CMA accepted that Anglian did incur increasing base costs to maintain the frontier level of leakage, which were not reflected in base models or the leakage enhancement allowance. The recognition of this more explicit cost-service relationship ultimately resulted in a £42.6m uplift to Anglian's base costs.
    - We encourage Ofwat to draw more explicit cost-service relationship for leakage through granular cost information for each main types of leakage reduction solutions.

[37] Ofwat service delivery report 2020-21 data.

[38] Paragraph 4.221 Anglian Water Services Limited, Bristol Water plc, Northumbrian Water Limited and Yorkshire Water Services Limited price determinations, Competition and Markets Authority.

Q22: Do you consider it would be feasible to assess the 'efficient' baseline performance level for each company for individual PCs such as leakage and PCC through econometric modelling? Are there any other PCs where you consider this could feasibly be attempted?

The consultation paper explores the possibility of using econometric modelling to determine an appropriate 2024-25 baseline performance level and to assess how a company has performed against an expected 'efficient' performance level over the historical period'. The econometric modelling will consider the influence of exogenous factors, endogenous factors and differences in historical levels of enhancement expenditure on company performance.

For the purpose of determining an appropriate 2024-25 baseline performance level, we think an adjustment mechanism as suggested in our response to question Q21 would be more pragmatic. The adjustment mechanism will help to refine 2024-25 'year 0' baseline performance level, as Ofwat can obtain updated forecast on 2024-25 performances through companies' PR24 business plan submission in October 2023 and through 2023-24 annual performance reporting in July 2024.

The econometric modelling could be deployed to improve our understanding on how exogenous factors like population growth, weather (temperature, rainfall) affects some performances (e.g., supply interruption, leakage, sewer flooding), and whether the impacts are significantly different among regions to warrant a company specific adjustment on either cost or PCL. The model can use independent data source for exogenous drivers wherever available. The modelling result will help reveal regional differences and whether they are significant enough to warrant a company-specific adjustment on either performance commitment levels, or the cost allowance to achieve performance commitment levels. Recognising the uncertainties of exogenous factors in PR24 and beyond, the econometric modelling can also help to establish a risk distribution on performances arising from the high/low cases of exogenous factors. This performance risk distribution will inform ODI design and the balance of the overall risk and return.

Data availability and quality in performance drivers are the key enablers to develop such econometric modelling. We encourage Ofwat to work with companies to establish a set of key exogenous drivers for core common PCs.

It may be possible to find suitable data for measures such as leakage, bursts, or supply interruption. However, investment to improve one of these areas is likely to improve all areas and so it is likely that a composite measure of performance will be required to avoid cost allocation issues. We consider this further in response to Q24.

Some measures are not suitable for this econometric modelling approach as they are significantly affected by other non-quantitative factors. An example would be, PCC, which is largely outside of companies control and significantly affected by government policy in water labelling and buildings standards.

Another econometric approach: Leakage an example

We have noticed Ofwat is not attempting to use quality service drivers in the botex models due to the potential endogeneity associated with management control of the driver. However, if the drivers are appropriately measured reflecting and balancing the regulatory incentives or if the endogeneity issue is appropriately treated using the available econometric techniques like instrumental variables within a panel dataset framework the endogeneity issue could be controlled.

The following analysis is just an *illustrative example* of what could be explored when adding some service drivers to the botex models. A more collaborative analysis developed jointly with the industry and third parties could bring robust models for benchmarking costs.

We have developed an example within the current econometric models using one of the most material drivers, leakage, in the treated water distribution model (TWD). We believe that leakage could be considered within an econometric model. We have measured the impact of leakage as the ratio between leakage and the total water produced by the company or the distribution input, in other words as it is presented in expression (6)[39]:

$$Leakage_{Rate}(\%)_{it} = \frac{Leakage_{it}(Ml/d)}{Distribution\ Input_{it}(Ml/d)} x100\% \qquad \textbf{(6)}$$

Using the leakage rate would give us the percentage of how much leakage is considered as a negative outcome in the sense of how much output is wasted through leakage. We should expect a negative impact on the effect of this service quality driver, which would reflect the expected efficiency incentive to reduce leakage from the regulatory point of view.

Table 13 – Treated Water Distribution (TWD)

Treated Water Distribution (TWD) with Leakage under Different Sample Size and Excluding Most Efficient Companies. TWDL=TWD1 with Leakage. The(year>) is the Sample Time period

	TWD1_CMA b/se	TWDL b/se	TWDL_(2012>) b/se	TWDL_(2013>) b/se	TWDL_(2015>) b/se	TWDL(NoPRT) b/se	TWDL(NoSRN) b/se	TWDL(NoSWB) b/se	TWDL(NoEFF) b/se
Ln(Mains)	1.055***	1.130***	1.129***	1.130***	1.161***	1.111***	1.145***	1.134***	1.132***
	(0.040)	(0.047)	(0.048)	(0.050)	(0.047)	(0.054)	(0.045)	(0.044)	(0.049)
Booster_per_Main_Km	0.570***	0.740***	0.741***	0.702***	0.678***	0.755***	0.881***	0.720***	0.901***
	(0.150)	(0.184)	(0.186)	(0.184)	(0.175)	(0.189)	(0.163)	(0.184)	(0.131)
Ln(Weighted_Density)	-3.338***	-3.961***	-3.962***	-3.889***	-4.323***	-3.905***	-3.799***	-3.750***	-3.444***
	(0.447)	(0.433)	(0.471)	(0.524)	(0.628)	(0.444)	(0.429)	(0.426)	(0.411)
Ln(Weighted_Dsty)^2	0.266***	0.314***	0.315***	0.309***	0.338***	0.312***	0.306***	0.298***	0.282***
	(0.031)	(0.030)	(0.033)	(0.036)	(0.044)	(0.031)	(0.031)	(0.030)	(0.029)
(Leakage/DI)		-0.033**	-0.034**	-0.034*	-0.036**	-0.036**	-0.044***	-0.034**	-0.051***
		(0.014)	(0.017)	(0.018)	(0.017)	(0.018)	(0.013)	(0.015)	(0.013)
constant	6.782***	9.385***	9.390***	8.989***	10.278***	9.415***	9.326***	8.609***	8.380***
	(1.510)	(1.611)	(1.737)	(1.959)	(2.305)	(1.610)	(1.666)	(1.594)	(1.541)
R2_Overall	0.962	0.963	0.961	0.960	0.964	0.960	0.971	0.965	0.975
Wald_Chi2	1745.518	1503.611	1658.562	1943.063	2255.028	1541.134	1707.525	1717.623	2414.696
RESET_P_value	0.05	0.32	0.33	0.32	0.34	0.38	0.40	0.27	0.40
BPagan_Test_P_value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Serial_Corr_P_Value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
StdDev_Ind_Effect	0.17	0.18	0.19	0.19	0.19	0.18	0.17	0.17	0.13
StdDev_Idiosy_error	0.14	0.13	0.13	0.12	0.10	0.14	0.13	0.13	0.13
Corr_comp_error	0.62	0.65	0.69	0.71	0.80	0.62	0.64	0.62	0.49
N_Sample_Size	21.00	21.00	21.00	21.00	21.00	20.00	20.00	20.00	18.00
T_Sample_Size	9.00	9.00	8.00	7.00	5.00	9.00	9.00	9.00	9.00
Observations	158.00	154.00	136.00	118.00	82.00	145.00	145.00	152.00	134.00

Source: Economic Regulation, Thames Water. Note: A model laballed as TWDL(NoSRN) means company Southern Water is not included.
* p<0.10, ** p<0.05, *** p<0.01

For example, if a company increases its levels of leakage and holding constant the distribution input, this will increase the leakage ratio. This situation could produce an incentive on leakage so the ratio would be incentivised to be reduced to its minimum producing an incentive to minimise botex *inefficient* expenditure. It could be argued that companies would have the perverse incentive to extract and produce more distribution input to keep the ratio holding, but this would increase costs significantly further showing inefficiency. Moreover, the extraction of water from the different natural sources is limited and controlled by the Environmental Agency thought

[39] We have used the cost assessment data on leakage and distribution input.

abstraction licenses, so this it is already restricted. We know that distribution input has to equal demand for water for customers and losses (leakage) where both have already the incentives to be minimised, one through water efficiency consumption and the other through leakage penalties, for instance.

The models presented in the Table 13 "*Treated Water Distribution (TWD)*" show the current CMA model TWD1_CMA in column one, followed by the same model with the inclusion of leakage rate, model TWDL[40]. In this TWDL model we can appreciate how the p-value of the functional form increases and also, we can notice the statistical level of significance that the leakage driver has on botex and the expected negative sign. The following models in Table 13 are a robustness check of the same model under different sample periods and sample sizes excluding the most efficient companies as a standard way to assess the robustness of the model. For instance, the model TWDL_(2016>) includes data from 2015-16 to 2019-20 and model TWDL_(NoEFF) excludes the most efficient companies from the sample, PRT, SRN and SWB. The consistency across the models is quite robust in all aspects. The magnitude of the scale coefficient across all models is greater than one suggesting diseconomies of scale.

Table 14 – Treated Water Distribution (TWD) Efficiency Ranks

Treated Water Distribution Efficiency Scores (2016-2020)				
Rank	**Efficiency score Current CMA-Ofwat Model**		**Model including Leakage**	
1	SWB	0.78	SWB	0.75
2	PRT	0.79	SRN	0.76
3	SRN	0.84	PRT	0.80
4	WSX	0.92	WSX	0.89
5	NWT	0.93	AFW	1.00
6	SVT	0.97	TMS	1.00
7	TMS	1.01	SVT	1.00
8	SES	1.04	SES	1.01
9	DVW	1.07	NWT	1.08
10	YKY	1.07	NES	1.09
11	AFW	1.09	WSH	1.10
12	WSH	1.10	DVW	1.10
13	NES	1.15	BRL	1.11
14	SSC	1.20	YKY	1.15
15	BRL	1.22	SEW	1.21
16	SEW	1.30	ANH	1.28
17	ANH	1.39	SSC	1.35
	UQ	**0.926**		**0.999**
	Mean	**1.05**		**1.04**

Source: Economic Regulation, Thames Water

Regarding the efficiency scores in Table 14 "*TWD Efficiency Ranks*" the top four positions of most efficient companies remain with the same set of companies, with some marginal changes in positions. This result is to provide a bit more of evidence on the implications of including leakage in the models and its impact on efficiency ranks across the industry[41].

[40] There are four observations less compared to the 158 observations in the TWD1_CMA model. This is because there is no reported data for leakage or Distribution input for SWB, SVT and DVW for year 2018-19 or 2019-20.

[41] It could be argued that the leakage rate still presents some endogeneity issues as it could be correlated with the time-invariant unobserved heterogeneity or the random error or a mixture of the two that could be related to (in)efficiency which might reflect the issue of management control as it was explained in our answer to question Q1. Using techniques such as Instrumental Variables could help to mitigate this. In panels, we could rely on the lagged

Q23: The need to collect further granular data to elucidate the cost-service relationship was highlighted by companies in response to our PR24 May consultation. Can you propose any data it would be proportionate to collect to support the high-level approach outlined in this chapter?

We welcome the initiative to collect more granular data for leakage in future annual performance reporting, this will help develop more explicit cost-service relationship for leakage. Useful information to collect include:

- Granular cost data for each of the main types of leakage reduction solutions.
- Key exogenous factors and endogenous factors.

Q24: What are your views on attempting to use of a composite variable to investigate the cost-service relationship, in the context of the methodological issues and complexities we outlined?

Considering that the investment for leakage, supply interruptions and bursts are highly correlated, for example, pressure management will benefit all three areas, there might be opportunity to use the three as a composite variable to investigate the cost-service relationship. Using a closely related composite variable would also reduce the complexity of cost allocation. The composite variable or Quality Index can be useful as this measure can mitigate data reporting issues or inconsistencies.

The management control perverse incentive or endogeneity of this driver will be relevant to how this Index will be used. The Index can be calculated using methods such as Principal Components Analysis (PCA) that will allocate the appropriate weights to the different variables in the Index based on the variability of the drivers. In other words, this weight is based on the variation across the principal components used in the calculation. The method assigns more weight to the component that has more variation. The following paragraphs are an example on how this PCA could be potentially implemented.

Principal Components Analysis (PCA), An Index Service-Quality Example

In this section ***we want to present an example*** of how the principal components could be used in the development of an Index of Service-Quality. With this example, we want to establish a clear link between costs and service-quality. We think that the current approach proposed in section 6 of the consultation is not clear with this respect and that there are different assumptions needed to make the proposal feasible. We think that the PCA approach could help to mitigate some of the concerning issues regarding the multidimensional view of services and also in its incorporation within the botex models establishing a clear link between cost and service. We also propose an econometric alternative and an example on how to deal with the endogeneity that the index could bring to the models if incorporated as an explanatory variable.

variables of the endogenous drivers as potential instruments to explore and to check its validity through the appropriate statistical test such as the Sargan-Hansen Test. Other external instruments could be developed with the regulator and the industry, where a good instrument is defined as a variable that is correlated with leakage but not with base costs. There are still some questions to tackle and procedures to check before we can get a robust approach. Perhaps the construction of an instrument could be more practical than the development of several performance commitments econometric models that are challenging to model and to develop.

We use the information on leakage, PCC and supply interruptions to develop the index. PCA is a technique that helps us to reduce the dimension of the data for a set of correlated variables. PCA allow us to reduce the dimension of these set of correlated variables (e.g., an index) to a dimension that collectively explain most of the variation of the original set of variables. The main idea of PCA is to reduce the number of drivers preserving as much information as possible from the set of drivers used in the analysis.

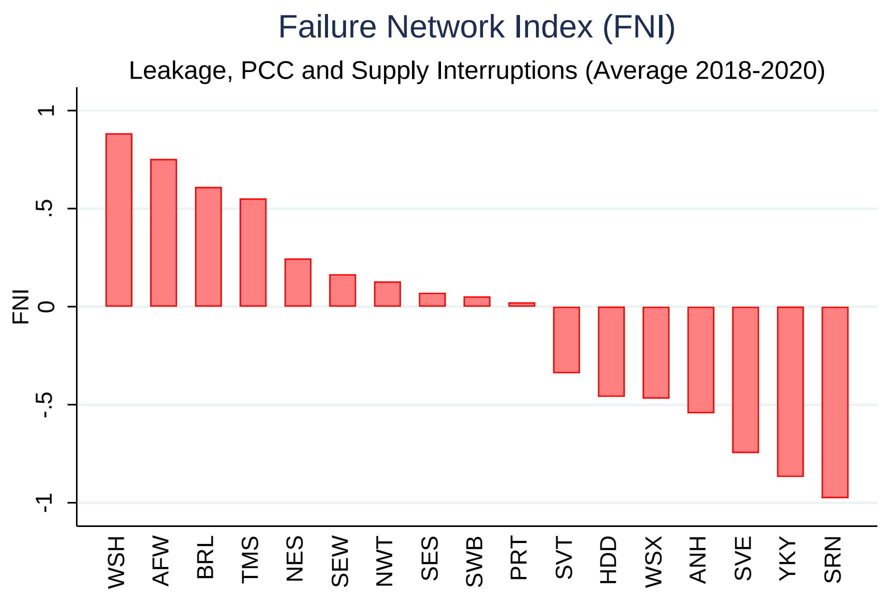
Principal components are new "variables or indexes" that are constructed as linear combinations of the initial set of drivers used in the analysis. Each of these principal components contribute to the explanation of the variation of the data, and the PCA identify the proportion of variation and the weight that each of the principal components has. As an illustrative example, the first principal component of the set of variables explored in this example look like the following expression (7):

$$PC_{1,it} = \phi_{11}(L_{it} - \bar{L}) + \phi_{21}(PCC_{it} - \overline{PCC}) + \phi_{31}(SI_{it} - \overline{SI}) \qquad (7)$$

Where, L is leakage, PCC is per capita consumption and SI is supply interruptions. The variables with the var on top represent the mean. The PCA has the restriction on the linear combinations or principal components on the form $\phi_{11} + \phi_{21} + \phi_{31} = 1$. The reason for doing this, is because if it is not, we could increase these ϕ's arbitrary in order to blow up the variance. The mathematical procedure is slightly more complicated, but we believe this brief explanation provides the basic intuition[42]. Figure 5 depicts the results of the principal components or linear combinations, in other words the Index that compiles these three variables as a composite Index with the corresponding weight derived from the analysis and framework of PCA. Please notice that the following results are ***only illustrative*** and that different assumptions can be made with respect to the data used, and the variables included in this index example. We have called this Index, Failure Network Index (FNI), as the improvements in the Index should reflect a reduction in each of the variables Leakage, PCC and Supply Interruptions.

[42] For more detail on how PCA works see James, Witten, Hastie and Tibshirani (2013). An Introduction to Statistical Learning. *Springer Texts in Statistics*.
PCA is easily implemented in Stata with the command pca. All variables were standardised for an appropriate comparison in PCA.

Figure 5 – Service-Quality Index using PCA



Source: Economic Regulation, Thames Water. Note: Principal Components Analysis.

The chart shows a set of companies with a positive FNI which means that these companies overall are above zero, reflecting the potential they still have to reduce more their levels in the Index to the margins of the companies with a neutral or negative index. This index is only based on data for the period 2018-2020. PCA produces an index for each company yearly which allow us to introduce it within the base econometric models as a clear way to represent the link between cost and service. Adding this Index FNI, into the Treated Water Distribution model provides the following econometric results presented in Table 15 below:

Table 15 – Treated Water Distribution (TWD) and Service Index

Endogenous Service Index: Treated Water Distribution Models

	TWD_CMA b/se	TWD(2017>) b/se	TWD_Endog b/se	TWD_Index_IV b/se
Lenghts of main	1.055***	1.088***	1.097***	1.101***
	(0.040)	(0.042)	(0.051)	(0.063)
Ln(BoosterPmpStn)	0.570***	0.433**	0.560**	0.514*
	(0.150)	(0.177)	(0.236)	(0.289)
Ln(Weighted_Density)	-3.338***	-3.130***	-3.200***	-3.054***
	(0.447)	(0.623)	(0.803)	(0.755)
Ln(Weighted_Dsty)^2	0.266***	0.245***	0.254***	0.243***
	(0.031)	(0.043)	(0.053)	(0.051)
Index_FNI			-0.107*	-0.071
			(0.065)	(0.156)
constant	6.782***	5.583***	6.070**	5.361*
	(1.510)	(2.043)	(2.424)	(2.890)
R2_Overall	0.962	0.965	0.951	0.957
Wald_Chi2	1745.518	1942.333	1378.046	561.941
RESET_P_value	0.05	0.05	0.61	0.69
StdDev_Ind_Effect	0.17	0.19	0.20	0.19
StdDev_Idiosy_error	0.14	0.09	0.07	0.08
Corr_comp_error	0.62	0.82	0.89	0.86
SarganHansenP_value				0.11
N_Sample_Size	21.00	19.00	17.00	15.00
T_Sample_Size	9.00	3.00	3.00	3.00
Observations	158.00	51.00	46.00	42.00

Source: Economic Regulation, Thames Water. Note: Index_FNI = Failure Network Index.
* p<0.10, ** p<0.05, *** p<0.01

The first column in Table 15 shows the CMA final econometric model for Treated Water Distribution, whereas the second column is the same model but based on data from 2018 to 2020. The reason for doing the analysis over this short period of 3 years in the second column of the table is because the FNI is calculated over this period as we do not have more information at the moment to build an Index based on more data to make it consistent with the whole period 2012-2020 of the botex models.

In this context, the third column (TWD_Endog) has the FNI in the model with a statistically significant effect and the expected negative coefficient for this "service-quality" index, as it is similarly explained in our previous example of leakage in question Q21. Although these results suggest the right direction of what could be expected from the Index and its effect on costs, it can be argued that the driver is still endogenous or that its coefficient and the other driver estimations could be biased. We understand that this endogeneity issue from the econometric point of view is related to the correlation between some part of the composite error term and the Index as explained in question Q1. We would like to provide a more detail explanation in what is, in our opinion the endogeneity problem before we provide a potential practical solution.

In principle, the inclusion of the quality dimension of the output produced by water companies is fundamental in assessing efficient costs. If excluded, we end up in a model (as they are currently established) that econometrically speaking faces omitted variable bias. For example, if leakage is excluded in a model of water distribution, this variable is highly correlated with length of mains, causing an endogeneity issue driven by omitted variables. This omitted variable would have implications on the assessment of economies of scale, efficiency scores and cost allowances calculations. Hence, a solution is to include a service quality driver in the base models (e.g., leakage, or an index that compiles several related drivers)[43].

However, the inclusion of a service driver in the base cost models, could cause an endogenous problem that is driven by the system or by the management control that the variable has as we explained in question Q1. This issue can cause econometric endogeneity problems. This economic regulatory incentive endogeneity issue could be translated into an econometric endogeneity problem. This is important to understand if we would like to mitigate its effect and provide a potential solution to solve the problem to some degree. Econometrically speaking, we think that the economic endogeneity characteristic of the driver service in the econometric RE models could come from two sources: *i*) correlation of the service driver with the unobserved heterogeneity of the random effects model or some part of the random noise as we explained in Q1, or through *ii*) a simultaneity issue of the structure of the cost model system between costs and service.

The first source of endogeneity *i*, could be tackled by the implementation of instrumental variables econometric techniques[44]. Using the lagged version of the endogenous driver FNI could be an alternative to find a valid instrument in the model. This is one of the advantages we have when using panel data models. Given the time dimension of the panel in water is relatively long with more than 10 years, this circumstance could be exploited to develop instruments. Unfortunately,

[43] We should recognise that a performance level of a service is not 100% base costs or 100% enhancement costs and that there could be a mix of the two. In some cases, the service (e.g., leakage, supply interruptions etc.) might be highly related to base costs rather than enhancement costs, therefore including services that are more related to base costs would be ideal.
[44] For a panel data model, this could be applied in Stata with the xtivreg command with the appropriate instruments.

in the current example this is not the case due to the short period we have for these service drivers, i.e., 2018-2020. If this alternative is not robust, we could think as an industry of potential instruments that could be built in collaboration with companies, and the regulator.

The second source of endogeneity could be more of a challenge to tackle than the first one, when we try to model quality or service as an independent model. The following expressions try to explain the problem of simultaneity in more detail. Let's assume a very simple example where Costs (C) and Service-Quality (SQ) are interrelated. Let's represent this interrelation in the following two simple expressions (8,9):

$$C = \gamma_1 + \beta_1(SQ) + u_1 \qquad \text{(Cost)} \qquad (\mathbf{8})$$

$$SQ = C + z_1 \qquad \text{(Service-Quality)} \qquad (\mathbf{9})$$

Where u_1 and z_1 are the errors of each model (e.g., it could represent the composite error of the random effects model). The simultaneity issue argues that the variables Cost, and Service are interdependent or jointly determined. Basically, this means that when companies or managers are deciding a particular improvement in leakage for example, this improvement has a direct effect on costs, and therefore, costs are explained directly by the decisions made to improve leakage. This interdependence could be seen more clearly if we substitute the previous (Service-Quality (9)) expression into the (Cost (8)) expression as follows:

$$C = \gamma_1 + \beta_1(C + z_1) + u_1$$

$$C = \gamma_1 + \beta_1 C + \beta_1 z_1 + u_1$$

$$C(1 - \beta_1) = \gamma_1 + \beta_1 z_1 + u_1$$

$$C = \frac{\gamma_1}{1-\beta_1} + \frac{\beta_1}{1-\beta_1} z_1 + \frac{1}{1-\beta_1} u_1 \qquad (\mathbf{10})$$

After some simple substitutions and algebra, the last expression (10) shows how costs depends not only on the error term of costs but also on the error term of (Service-Quality) z_1 in equilibrium. Similarly, the equilibrium expression (11) for Service-Quality in this example is:

$$SQ = \frac{\gamma_1}{1 - \beta_1} + \frac{1}{1 - \beta_1} z_1 + \frac{1}{1 - \beta_1} u_1 \qquad (\mathbf{11})$$

This expression tells us that any external induced change in z_1 will cause an effect on costs and service-quality simultaneously. This could be represented to decisions involved in the unobserved heterogeneity term of the random effects model for example, that could be reflect some (in)efficiency capabilities of managers in taking optimal decisions. Therefore, modelling a separate performance level with previous expenditures could bring the simultaneity endogeneity issue in the analysis too. This is a more difficult problem to tackle, and the solution that sits in Simultaneous Equation Models (SME) could be really difficult to implement in a panel data framework given the time scales and technical challenges that this methodology brings. For this reason, we opt to tackle the endogeneity problem from the source *i* as explained above. We think that tackling the problem from this angle is practical and could bring another alternative to consider with the one already proposed in the consultation in section 6.

Within this context and using the small time period of only three years due to the limited availability of data of the service drivers used in this example, we have used the lagged (t-2) of the Index FNI as a potential instrument variable to use in the model to mitigate the potential endogeneity problem described in this question and Q1[45]. As it is appreciated from this ***illustrative example***, the last column TMD_Index_IV in the Table 15 shows that the coefficient of Index_FNI goes from -0.107 to -0.071. This might be an indication of the potential bias that the driver has in the initial model TWD_Endog. Moreover, the other cost drivers in the model are also changed. It is interested how all the estimated coefficients change when the service is included versus the situation when it is not included. This is easily seen across all the econometric results in the table. Unfortunately, in this example, the coefficient of Index_FNI after controlling for its potential endogeneity is not significant. This might be explained by the small sample time period used but the endogeneity issue has been controlled as there is a significant change in the coefficient, as well as in the rest of the drivers of the model. In the last model, we have included the P-value for the Sargan-Hansen Test for the instrument variable validity used in this example. Moreover, with this kind of models we could explore the cost of providing drinking water and service-quality. This could be quantified by including an interaction term with the service-quality variable and its respective derivative with respect to cost complementarities between the two products, in this case Mains and Quality such as $\frac{\partial^2 C}{\partial Mains\ \partial FNI\_Index} \leq 0$, for instance. Ideally, would be more intuitive if we have water delivered and quality as the two outputs produced to derive the cost complementarities.

To sum-up, with this example ***we illustrate*** the potential of PCA and the econometric techniques interaction that are available in the econometric toolbox that could help us to deal with different challenges between the cost-service link and the benchmarking modelling framework. If implemented with the right data and collaboration of the industry, we could have an alternative to the proposal of dealing with the cost-service relationship presented in section 6 of the consultation. In this way we can have a ***clear link between costs and service*** directly specified and with the endogeneity issue mitigated.

Q25: Do you have any proposals for how to make adjustments where a performance commitment level differs from that expected to be delivered from base costs?

As mentioned in our response to question Q22, we encourage Ofwat to work with companies to establish a set of key exogenous drivers for core common PCs and develop econometric modelling to reveal regional differences and assess whether they are significant enough to warrant a company-specific adjustment on either performance commitment levels, or the cost allowance to achieve performance commitment levels.

PR24 ODI rates for service and environmental PCs will be determined through the collaborative customer research, rather than a combination of marginal cost and benefit as applied in the past price reviews. This means the ODI rates will not necessarily reflect the efficient cost required to deliver performance improvement. For this reason, we consider the enhancement investment on performance should be assessed and funded through enhancement cost allowances rather than through ODI payments as ODI rates will not take into account or reflect efficient marginal cost.

[45] Another potential example of an instrument could be the rapid temperature changes or its variation.

7. Residential Retail Cost Assessment

Q26: Do you have any comments regarding our proposal to ask companies to separate out the part of their provision of bad debt costs to do with Covid-19 that was made outside of their standard methodology in the PR24 business plan tables?

We understand the importance of using reliable data, and Ofwat's desire to separate out provisioning for Covid, and in principle would agree that this should be done. We made a provision for future losses in respect of Covid at 31st March 2020 and 2021, which have been audited by PwC, and for which the methodology can be shared.

Clearly judgements were required to be made as part of the forecasting process regarding the expected impact of Covid. We used econometric forecasts based on employment data in the London area, as part of our process. We assessed the Covid impact to be 3.25% of budgeted cash flow i.e., c. £50m using this econometric data. Although anecdotally the actual cash flows would support this assessment, existing reports do not enable this to be verified. We do not currently have a reporting process able to measure this impact on cash flows.

This process will also offer opportunity for companies to seek to maximise the impact of Covid, which could mask underlying underperformance. It will require high level of clarity/transparency to ensure fair reporting takes place.

Q27: What guidance would aid companies to provide appropriate data related to the provision of bad debt costs to do with Covid-19?

Clear industry KPIs and methodology would be helpful to ensure a common approach. This should include:

- Clarity as to whether its Covid + economic impact associated with Covid, or just Covid.
- Clear criteria for what to include as 'Covid related' and what not to include as there are significant other items affecting bad debt costs including Brexit, energy price increases, changes to universal credit etc. Some of these will have an equal or bigger impact on our customers and therefore the costs for collections and the debt risk.
- Assessment of general economic climate and impact on affordability. For instance, Auditors PwC have estimated that the cost of living for a typical earner c. £30k, has increased by 7%.
- Clarity on how Covid may have impacted customer behaviour (and whether the changes have had a negative or positive effect) e.g.
    - Vulnerability and affordability.
    - Extended payment terms or non-payment.
    - Impact on transience within city centres.
    - Movement of population away from city centres and rented accommodation.
    - Consumption patterns.

We are concerned that the most significant impact has already occurred during the previous two years without appropriate monitoring/measuring systems in place. We have schemes such as Breathing Space to support vulnerable people. It would be good to understand the use of such schemes by others and the use and support of these need to be captured.