Thames Water

Response to Econometric Base Cost Models for PR24 Consultation

Wholesale Water Questions

Q3.1) Do you agree with our proposed set of wholesale water base cost models?

There are aspects of your proposals that we agree with and aspects that we do not agree with. We set our views in more detail in response to the more specific questions.

We welcome the inclusion of Average Pumping Head (APH) and the more disaggregated density driver captured by MSOA density. We think that the inclusion of these drivers is an improvement with respect to the PR19 models. However, we think that there are economic, econometric, and engineering reasons to disregard some of the proposed models across TWD and WW.

Given this, we neither agree nor disagree with the current suit of proposed models for PR24.

Q3.2) Do you agree with the inclusion of average pumping head in a sub-set of treated water distribution and wholesale water models?

We strongly support the inclusion of APH in TWD and WW models. The variable has a strong engineering rationale (as recognised by Ofwat[1] and the CMA[2]), it is largely exogenous, and has strong econometric evidence to support it. The booster pumping stations measure is inferior to the APH in every respect and should be dropped.

Context

It has been recognised that topography can have material cost implications on water companies. Topography can determine the extent of pumping (energy) required to transport water from source to tap.

At PR19 Ofwat used the number of booster pumping stations per length of mains (BPSM) to capture these effects.

---

[1] See Ofwat CAWG on APH 07/07/2021, slide 9 on: https://www.ofwat.gov.uk/wpcontent/uploads/2021/10/CAWG_07.09.2021.pdf

[2] See CMA PR19 FD on: https://assets.publishing.service.gov.uk/media/60702370e90e076f5589bb8f/Final_Report_---_web_version_-_CMA.pdf , p. 139. Or "*Pumping water is energy intensive and APH is considered a proxy for the energy requirement of companies*", p. 132.

It is widely accepted that from an engineering and economic point of view the APH is a better cost driver than the BPSM. Compared to the BPSM, APH is significantly less under management control and significantly more related to pumping requirements forced by the region's topography.

While BPSM rather than APH was used at PR19, this was driven mainly by concerns with the quality of APH reporting. These concerns have been largely addressed, in particular within the treated water distribution service.[3]

With new years of available data since PR19 and with improvement in the quality of reported data, the APH is now statistically significant and stable in all the models. This further validates the improvements in data quality.

The engineering, economic and econometric evidence in support of the APH over the BPSM are compelling. In light of that and the further concerns with the BPSM, which we raise below, we consider that there is a strong case to remove the BPSM from the models and instead use APH as the variable that explains energy consumption due to topographically forced pumping.

BPSM concerns

Below we raise a number of concerns with the BPSM variable in cost assessment models.

1. High correlation between density and BPSM

The BPSM is highly negatively correlated with the density measures. The APH, on the other hand, is not. This can be seen in table 1.

Table 1: correlation of BPSM and APH with density measures

	Ln(BPSM)	Ln(APH)
Ln(MSOA Density)	-0.60	-0.16
Ln(MSOA to LAD Denisty)	-0.67	-0.10
Ln(Property per length of mains)	-0.67	-0.18

Source: Economic Regulation, Thames Water

High correlation between explanatory variables is known to have negative implications on the quality of econometric models. It reduces the accuracy of the estimated

[3] See Ofwat consultation PR24 Econometric Consultation or Turner & Townsend, WRC (24 March 2022), Average Pumping Head: data quality improvement.

coefficients.[4] The high correlation between BPSM and (population or property) density suggests that both variables capture similar factors, making it difficult for the models to isolate the real effect of each driver (similar to Ofwat's argument in the case of Regional Wages versus Density).

Put differently, high correlation reduces the likelihood that the estimated coefficients are close to their true value. The estimation would be very sensitive to the sample at hand, or the expected sign of the driver may have the wrong sign or implausible magnitudes[5].

Table 2 illustrates the instability of the BPSM due to its correlation with other drivers (included or omitted), compared with the stability of the APH using all the years in the sample. The sign of BPSM changes from positive to negative when removing density from the model (likely due to an omitted variable issue), and it becomes significantly more negative when removing the scale variable as well. This does not happen with APH. The APH coefficient remains within a stable and narrow range (0.336 to 0.411) when removing the same variables.[6]

---

[4] The VIF test for detecting multicollinearity is just one tool among others and it cannot be used as the only one or definite test to conclude the absence of multicollinearity problems in a multivariate model, given that the VIF has its own weaknesses. In addition, the VIF >1 by definition, implies collinearity issues among the predictors. The "accepted" VIF thresholds of 4, 5 or 10, are just a "Rule of Thumb" without any theoretical background. This "rule" was proposed in an article by Marquardt, D. (1970) "*Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation*" in the journal Technometrics. This "rule" or suggestion made by the author is based on *an illustrative example* that the author makes in his article but without any generalization or theoretical background. See Marquardt, D. (1970), particularly between pages 605 to 610, where the illustrative example is presented. VIF should be interpreted with caution and complemented with other multicollinearity tools such as simple pair-wise correlations (as Ofwat explores), or Condition Number (See Greene (2012), Econometric Analysis, 7th edition, p. 130). VIF should be considered within the context is run, such as its sample size as proposed by O'Brien, R. (2007), "A Caution Regarding Rules of Thumb for Variance Inflation Factors", Journal of Quality & Quantity. This will lead to different interpretations, and the "rule" will probably not apply.

[5] There are generally three symptoms to suspect that there are multicollinearity issues in a model. Greene (2012, p. 129), *Econometric Analysis*, Seventh Edition, mention them:

1) Small changes in the data produce wide swings in the parameters estimates,
2) Coefficients may have very high standard errors and low significance levels even though they are jointly significant and the R^2 for the regression is quite high.
3) Coefficients may have the "wrong" sign or implausible magnitudes.

[6] We demonstrate using TWD2 and TWD5, but the same results are obtained with any TWD or WW model included in the consultation.

Table 2

Static Panel-Data Models: TWD Water PR24 Econometric Consultation Sample Time Period: 2011-12 to 2021-22

	TWD1-2 b/se	TWD2-2 b/se	TWD3-2 b/se	TWD2 b/se	TWD1-5 b/se	TWD2-5 b/se	TWD3-5 b/se	TWD5 b/se
Ln(BoosStn_per_Main)	-0.155	-0.028	0.555***	0.433***				
	(0.656)	(0.249)	(0.151)	(0.128)				
Ln(Mains)		1.058***	0.978***	1.026***		1.055***	0.961***	1.017***
		(0.050)	(0.054)	(0.050)		(0.052)	(0.044)	(0.019)
Ln(MSOA_Density)			0.712***	-5.561***			0.521***	-6.539***
			(0.181)	(1.316)			(0.145)	(0.928)
Ln(MSOA_Dsty)^2				0.393***				0.445***
				(0.080)				(0.058)
Ln(APH_TWD)					0.371	0.336**	0.379***	0.411***
					(0.239)	(0.139)	(0.110)	(0.051)
constant	3.634	-5.935***	-8.338***	15.638***	2.665**	-7.260***	-10.635***	16.573***
	(2.723)	(1.222)	(1.066)	(5.030)	(1.040)	(0.997)	(1.477)	(3.670)
R2_Overall	0.025	0.889	0.939	0.952	0.005	0.894	0.946	0.965
N_Sample_Size	19.000	19.000	19.000	19.000	19.000	19.000	19.000	19.000
T_Sample_Size	11.00	11.00	11.00	11.00	11.00	11.00	11.00	11.00
Observations	187.00	187.00	187.00	187.00	187.00	187.00	187.00	187.00

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

In table 3, we report a sensitivity analysis for model TWD2-2 (shown in table 2), where length of mains and BPSM are included. We do this initial analysis to understand the stability of the BPSM coefficient and compare it with the stability of the APH coefficient before density is added in the models. This provides information on how dependent BPSM and APH are to the inclusion of population density in a subsequent model. Table 3 shows the results for model TWD2-2, where the sign of the BPSM coefficient changes depending on the years included in the sample. For example, if we remove year 2022 in model TWD2-2_22, the BPSM coefficient is positive, whereas if we remove year 2012 in model TWD2-2_14, the coefficient becomes negative. This BPSM coefficient instability persists when other years are removed from the sample. Moreover, BPSM is not statistically significant in this specification.

Table 3

Model TWD2-2 (BPSM): Robustness Tests

	TWD2-2 b/se	TWD2-2_2022 b/se	TWD2-2_2021 b/se	TWD2-2_2020 b/se	TWD2-2_2012 b/se	TWD2-2_2013 b/se	TWD2-2_2014 b/se
Ln(Lenght_of_Mains)	1.058***	1.056***	1.054***	1.047***	1.053***	1.049***	1.048***
	(0.050)	(0.051)	(0.054)	(0.050)	(0.051)	(0.051)	(0.047)
Ln(BPSM)	-0.028	0.029	0.103	-0.023	-0.095	-0.146	-0.140
	(0.249)	(0.244)	(0.250)	(0.239)	(0.252)	(0.250)	(0.242)
constant	-5.935***	-5.682***	-5.355***	-5.841***	-6.166***	-6.348***	-6.306***
	(1.222)	(1.213)	(1.243)	(1.227)	(1.234)	(1.240)	(1.222)
R2_Overall	0.889	0.886	0.882	0.891	0.889	0.890	0.889
Wald_Chi2	479.932	466.000	417.400	513.562	463.948	463.522	585.069
RESET_P_value	0.50	0.50	0.50	0.49	0.48	0.46	0.46
BPagan_Test_P_value	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_Sample_Size	19.00	19.00	19.00	19.00	19.00	19.00	19.00
T_Sample_Size	11.00	10.00	9.00	8.00	10.00	9.00	8.00
Observations	187.00	170.00	153.00	136.00	170.00	153.00	136.00

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

We do the same analysis for APH using model TWD2-5 in table 2. The results are shown in table 4.

Table 4

Model TWD2-5 (APH): Robustness Tests

	TWD2-5 b/se	TWD2-5_2022 b/se	TWD2-_2021 b/se	TWD2-5_2020 b/se	TWD2-5_2012 b/se	TWD2-5_2013 b/se	TWD2-5_2014 b/se
Ln(Lenght_of_Mains)	1.055***	1.048***	1.041***	1.044***	1.055***	1.057***	1.056***
	(0.052)	(0.053)	(0.055)	(0.053)	(0.051)	(0.052)	(0.051)
Ln(APH_TWD)	0.336**	0.285*	0.300*	0.317**	0.348*	0.339*	0.280*
	(0.139)	(0.148)	(0.178)	(0.151)	(0.186)	(0.187)	(0.153)
constant	-7.260***	-6.976***	-6.962***	-7.096***	-7.313***	-7.295***	-7.017***
	(0.997)	(1.039)	(1.136)	(0.986)	(1.159)	(1.141)	(0.966)
R2_Overall	0.894	0.893	0.891	0.895	0.893	0.892	0.891
Wald_Chi2	844.905	767.070	479.967	459.507	677.164	581.517	530.706
RESET_P_value	0.48	0.44	0.45	0.47	0.49	0.49	0.44
BPagan_Test_P_value	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N_Sample_Size	19.00	19.00	19.00	19.00	19.00	19.00	19.00
T_Sample_Size	11.00	10.00	9.00	8.00	10.00	9.00	8.00
Observations	187.00	170.00	153.00	136.00	170.00	153.00	136.00

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

The APH coefficient in table 4 is robust and does not change when removing from the sample the same years as for BPSM in table 3. The robustness of APH is seen in the stable positive sign, its consistent significant statistical effect, and its magnitude ranging between [0.280 and 0.348] (in contrast to BPSM ranging between [-0.146 and 0.103]). We consider the results of the BPSM cost driver concerning. It can be argued that the model TWD2-2 with BPSM and length of mains might be affected by an omitted variable issue, such as the omission of population density due to its high correlation with BPSM. Only when population density is added in the model does the sign of the BPSM coefficient becomes stable. This is not the case for APH. APH coefficient is robust (and its sign stable) regardless of population density added or omitted from the model, as shown in table 2 and 4[7]. These results support the argument that while APH is fully

[7] We have also carried out a similar analysis as presented in Table 2 for the WW models. For this analysis using WW models there is some differences as the scale driver used is different, Properties. However, the same results are found as presented in Table 2. For instance, in the WW models, the pair-wise correlation between Ln(Properties) and Ln(BPSM) is -0.28, which is considerable high with respect to the correlation in the TWD models between the scale Ln(Mains) and Ln(BPSM) that is -0.11. Although the correlation -0.28, in principle, should not be of concern, a Condition Number analysis of collinearity reveals the contrary. We calculate the Condition Number Index for the models that uses Properties and BPSM only (equivalent to model TWD2-2, in table 2), and the results are:

	Condition Number Index	**Variance-Decomposition Proportions** _cons	Ln(Properties)	Ln(BPSM)
1	1	.	.	.
2	30.22	.	0.65	0.63
3	39.24	1	0.35	0.37

Source: Economic Regulation, Thames Water.

As suggested in Greene (2012, p. 130), values greater that 20 in the Condition Number are suggested as indicative of a collinearity problem. In this case, the second condition number of 30.22, suggests an issue. Moreover, the values of the variance-decomposition proportions are quite useful, as values greater than 50%, are indicative of the source of collinearity if more than two variables are involved. This information indicates that for the stability of the sign of BPSM in a context where properties is used as the scale factor, BPSM depends on the inclusion of properties to stabilise its sign due to his correlation as shown by the Condition number index. This result provides more information of the dependency of Properties or a scale factor, in the context of the WW models. in TWD it is more dependent on density, but overall, both influence BPSM stability. On the contrary, APH is not. For more information on how to run the test in Stata see the command, `coldiag2`.

exogenous in the models, BPSM is not, as it depends on the inclusion of population density.

To summarise, there are implications to the use of highly correlated explanatory variables: the estimated coefficients are not very accurate[8] and can vary significantly with the sample at hand or the inclusion/exclusion of other variables. This can have material implications to efficiency assessment and cost allowances for companies.

Ofwat has an alternative measure, the APH, which is uncorrelated with density (among its many advantages over the BPSM) and should be used instead of BPSM.

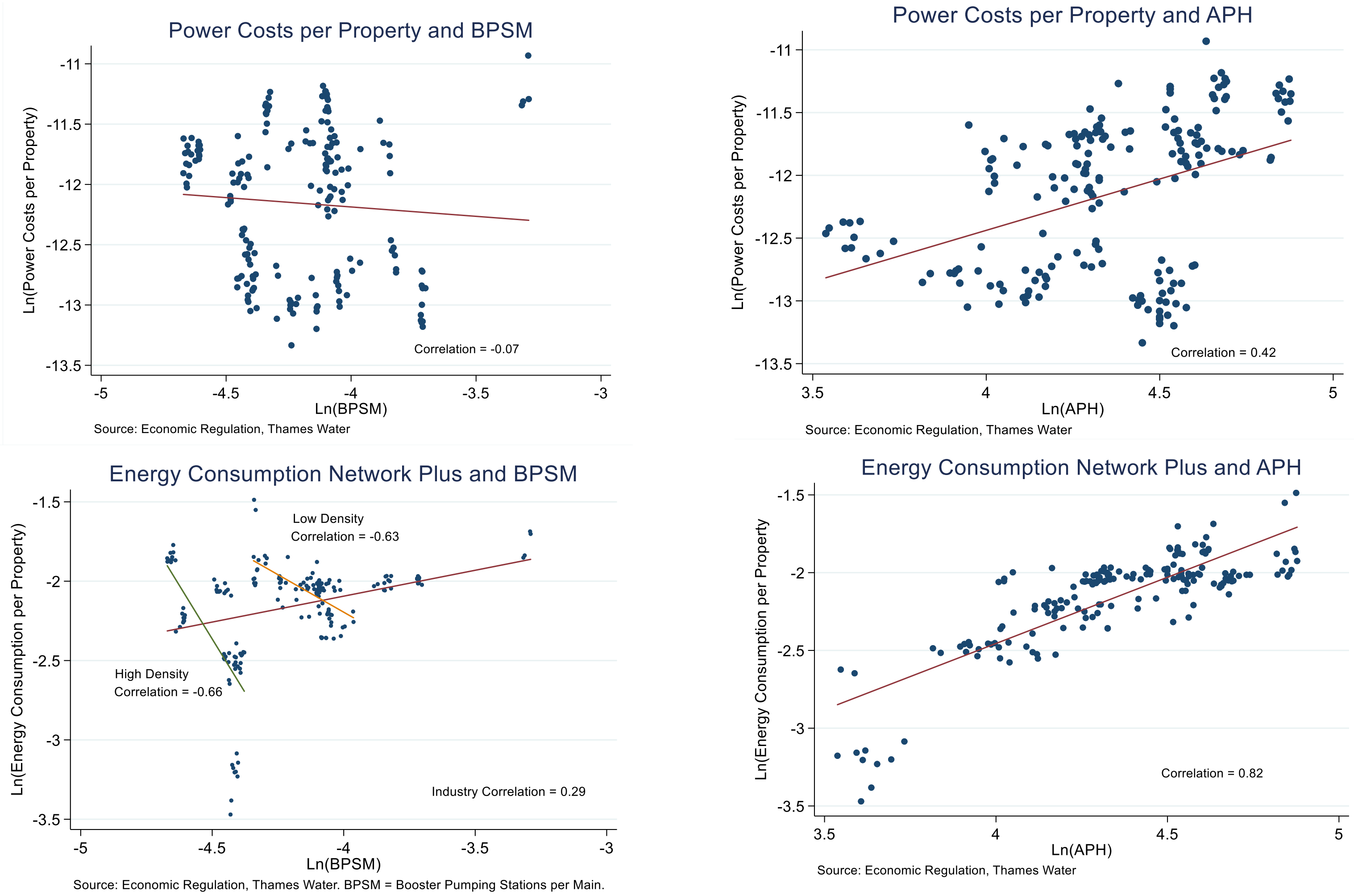
[8] One implication is on the estimated elasticity of density that we show in section 3 of this answer.

2. BPSM is not correlated with power costs or energy consumption

If APH and BPSM are meant to capture energy/pumping costs forced by regional topography, we would expect them to correlate with power costs (per property) and energy consumption (per property).

Figure 1 shows that this is the case for APH, but not for BPSM. The correlation of APH with power costs per property (0.42) or with energy consumption per property (0.82) is positive and high[9]. On the other hand, the correlation of BPSM with the same variables is a lot weaker. In fact, the correlation of BPSM with power costs per property is negative (-0.07)[10]. This is counter-intuitive and raises a serious concern that BPSM is not a relevant cost driver for topographical pumping costs.

Figure 1



[9] We correlate BPSM with a standardised power costs or energy consumption. If we were not standardising them by scale, then the correlation would be distorted – we do not expect a positive or negative correlation between BPSM and power costs, because power costs are primarily determined by scale. The scale effect must therefore be removed.

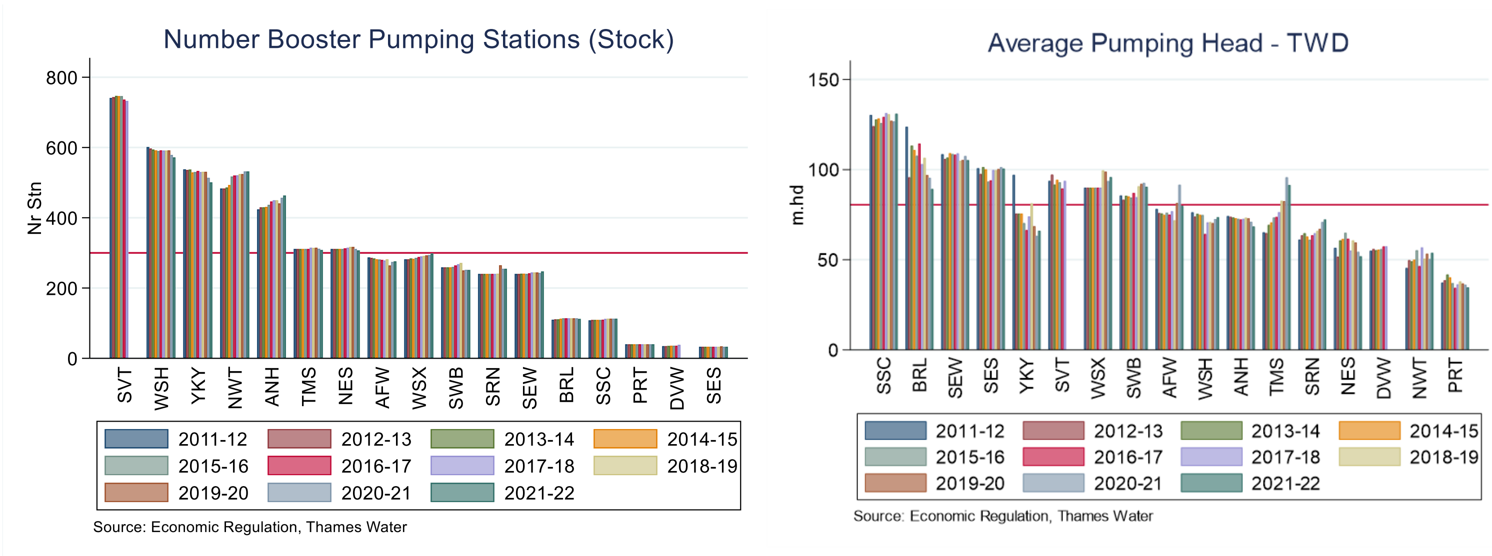
[10] We explain in more detail the relationship between BPSM and Energy consumption per property (0.29) in Figure 5.

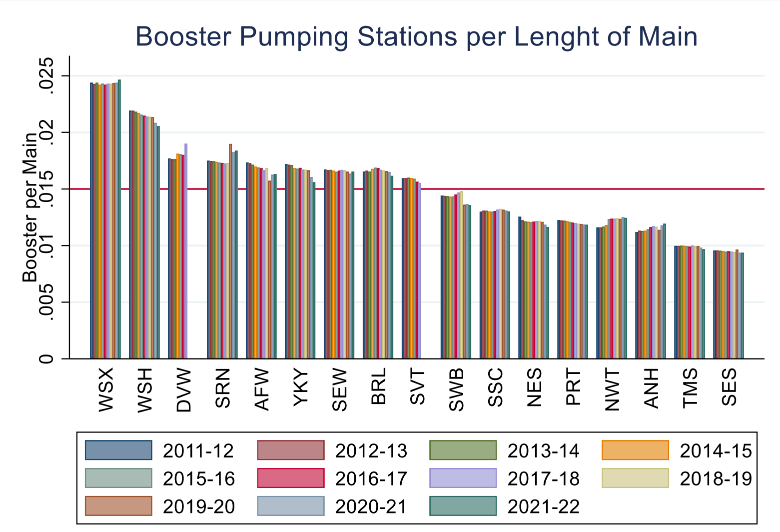
While the negative correlation of BPSM with power costs per property is counter-intuitive, if we expect BPSM to capture pumping cost requirements, it may not be counter-intuitive if we consider other factors that affect a company's decision on the stock of pumping stations.

For example, there are companies with the same number of pumping stations over the 11-year period of the dataset despite the fact that the scale of the business, approximated by the length of mains, has increased (thus making BPSM decrease) while for other companies the number of pumping stations increases in tandem with the network (thus BPSM remains stable).

The point is that the stock of pumping stations is more fixed in the short term compared to APH – see figure 2 that shows that APH is more variable year-on-year compared to the number of boosters pumping stations and BPSM. Once the new stock of stations meets the requirements of the expected demand, it tends to be stable and fixed. It is often more efficient to install a booster pumping station today with spare capacity to meet future demand growth, rather than to adjust the number of stations year on year. APH is more reflective of changes in the conditions of the current demand (e.g., seasonal conditions where more water needs to be pumped, for example, in hot summers). Thus, APH is better linked to the energy requirements of companies in any reporting year.

Figure 2

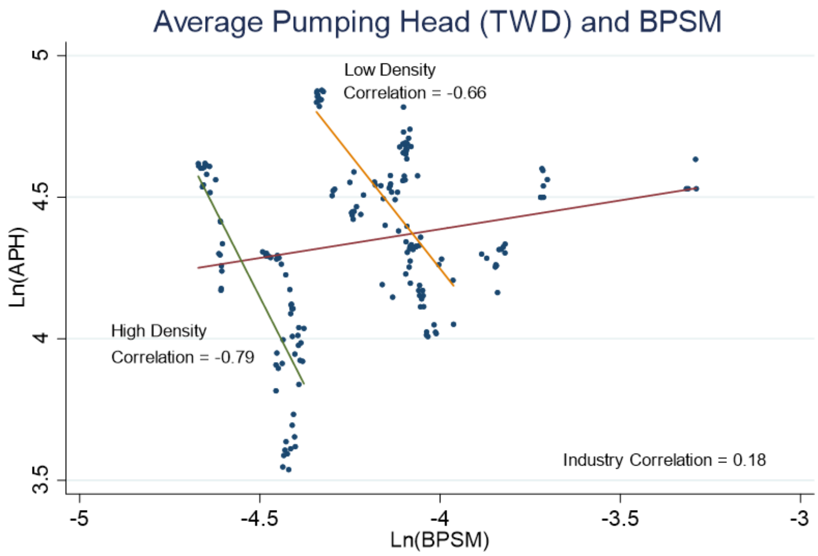




Source: Economic Regulation, Thames Water

Another illustrative way to explore what is BPSM actually capturing in the models is to consider its correlation with APH. If BPSM is capturing topography factors, we should expect a correlation with APH, given that APH has a strong engineering rationale as a measure of the energy intensity that is used by a company to pump water.

Figure 3



Source: Economic Regulation, Thames Water. BPSM = Booster Pumping Stations per Main.

Figure 3 shows that the correlation is weak (0.18). A closer look at the chart reveals two clear clusters of companies. The first cluster includes five of the denser companies in the industry (i.e., TMS, PRT, NWT, SES and NES) alongside with Anglian, which is large and sparse. The correlation of BPSM and APH within the cluster is negative, shown by the green line.

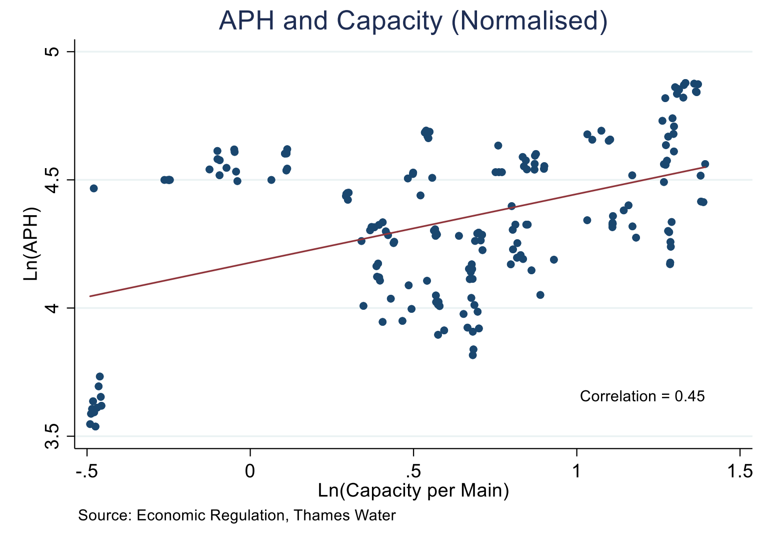
In the second cluster we find many companies with low density, except, AFW, BRL and SRN. In the second cluster, again, the correlation is negative, shown by the yellow line.

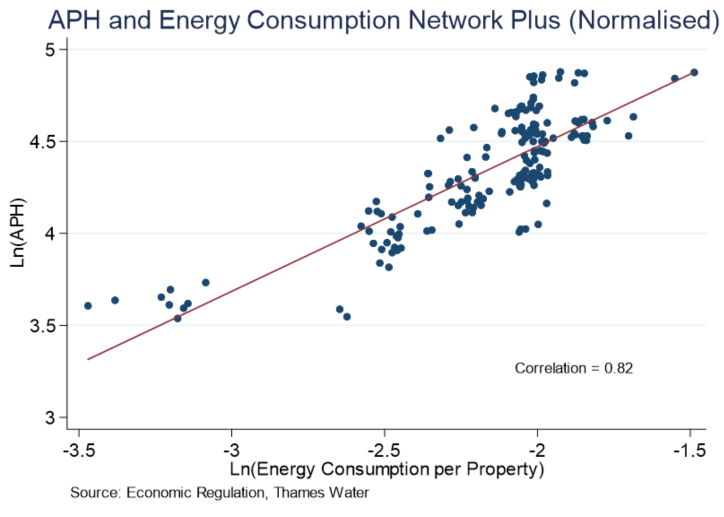
This negative correlation within each of the clusters and the positive correlation across the industry is known as the Simpson's paradox. The implication of the paradox is that it could lead to misinterpretation of the data at the industry level, for example to lead to an incorrect conclusion about the expected relationship of APH and BPMS. It might also highlight the importance of subgroups analysis and it also highlights the importance to proceed with caution when using statistical models regarding the driver.

Given the weak relationship between APH and BPSM, it is hard to argue that the two measures are, to some extent, substitutes.

We also explored the correlation of APH with total capacity of pumping stations per length of mains, which we consider a more appropriate driver of pumping requirements than the BPSM. We show that the correlation of APH and Capacity per Main is positive and around 0.45 corroborating a more technical relationship between APH and capacity than with BPSM. Also, the correlation with energy consumption per property and APH is robust, 0.82, corroborating the link that APH has with energy and capacity. These intuitive results can be seen in figure 4 (in addition the clustering issue or Simpson's paradox disappear).

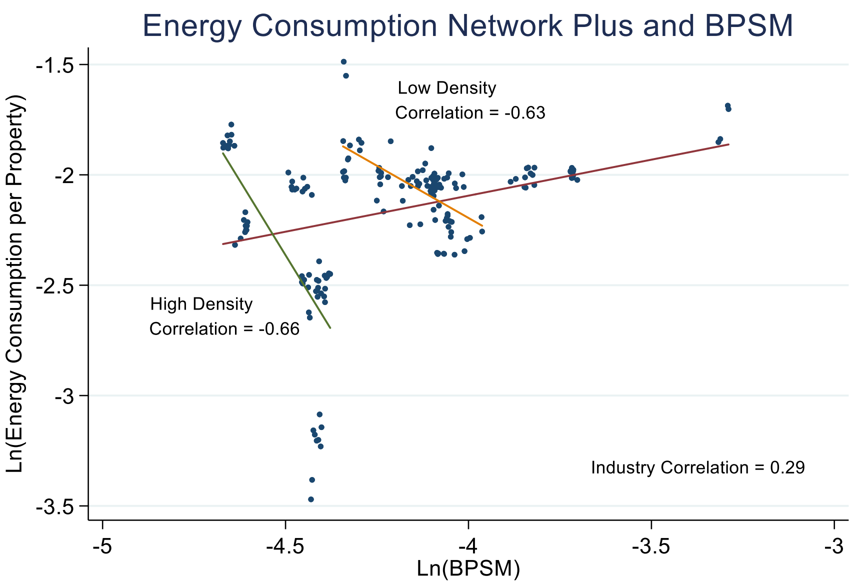
Figure 4





On the contrary, the Simpson's paradox does not disappear when correlated with energy consumption per property as it is presented in figure 5.

Figure 5

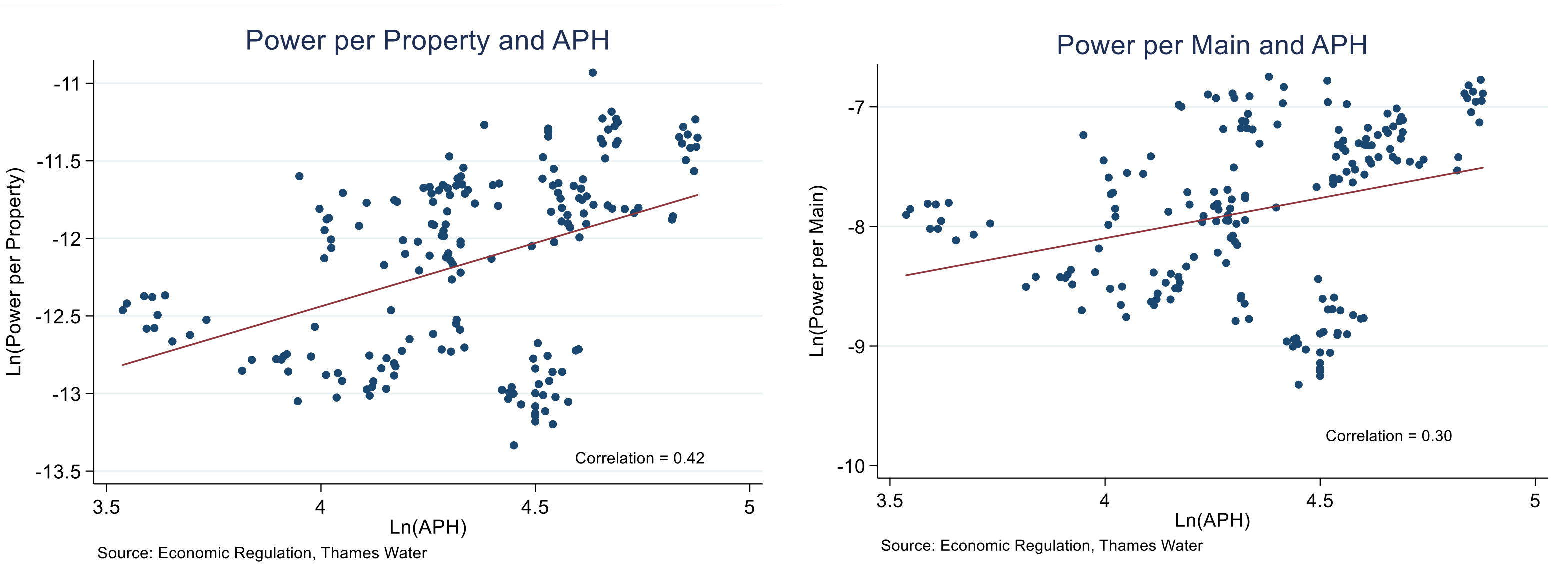


Source: Economic Regulation, Thames Water. BPSM = Booster Pumping Stations per Main.

Finally, when using APH the expected positive correlation with power costs is found at 0.42 with TWD power costs per property and 0.30, if power costs are normalised with length of mains. This is illustrated in figure 6[11].

[11] We have noticed that for APH and Botex cost normalised with property and Mains the correlations are positive and 0.46 and 0.24, respectively. This provides more evidence about the robustness of APH. On the contrary, when correlating BPSM with power costs per property (-0.07) or mains (-0.30) the correlation is negative. Moreover, when correlating BPSM with botex per property (0.38) and mains (-0.16) the result is not robust versus APH. This evidence shows once again the instability of BPSM due to its multiple interactions with other factors (e.g., scale).

Figure 6



We can conclude that there is strong intuitive and empirical evidence that suggests that BPMS or Number of Booster Pumping Stations (NPS, the stock) is capturing information related to the scale or complexity of the network alongside with density or spreading factors related to the operating area of the company. There is no information that suggests that BPSM tends to capture what it is argued to capture regarding topography or amount of work to pump water. There is also economic rationale and empirical evidence that suggests that APH does capture the intensity of energy used by companies and that it provides a stronger and coherent relationship with power costs as it is intended to capture. The next sections show the serious econometric implications that BPSM is adding in the models.

3. Sensibility of density coefficients across models: APH v BPSM

We are concerned about the plausibility of the density coefficients in models that include BPSM. Table 5 compares the estimated coefficient of TWD models without APH or BPSM with the coefficients estimated from the corresponding models in the consultation (i.e., models that include APH or BPSM).

The results show that the estimated density coefficients are sensitive to the inclusion of BPSM in the models. That is, when BPSM is included, the coefficient of density changes significantly. On the other hand, the coefficient of density is stable and robust to the inclusion or exclusion of APH. The inclusion of BPSM in the model has implications on the elasticity of density, which we explore in more detail in the next section.

Table 5

TWD Models PR24 Econometric Consultation with MSOA

	TWD_Excl b/se	TWD5 b/se	TWD2 b/se
Ln(Mains)	1.027***	1.017***	1.026***
	(0.049)	(0.019)	(0.050)
Ln(MSOA_Density)	-6.710***	-6.539***	-5.561***
	(1.547)	(0.928)	(1.316)
Ln(MSOA_Dsty)^2	0.455***	0.445***	0.393***
	(0.094)	(0.058)	(0.080)
Ln(APH_TWD)		0.411***	
		(0.051)	
Ln(BoosStn_per_Main)			0.433***
			(0.128)
constant	18.997***	16.573***	15.638***
	(6.216)	(3.670)	(5.030)
R2_Overall	0.947	0.965	0.952
N_Sample_Size	19.000	19.000	19.000
T_Sample_Size	11.00	11.00	11.00
Observations	187.00	187.00	187.00

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

TWD Models PR24 Econometric Consultation with LADtoMSOA

	TWD_Excl b/se	TWD4 b/se	TWD1 b/se
Ln(Mains)	1.070***	1.062***	1.070***
	(0.046)	(0.030)	(0.043)
Ln(MSOAtoLAD)	-3.062***	-2.975***	-2.729***
	(0.627)	(0.451)	(0.502)
Ln(MSOAtoLAD)^2	0.235***	0.229***	0.219***
	(0.043)	(0.031)	(0.035)
Ln(APH_TWD)		0.357***	
		(0.067)	
Ln(BoosStn_per_Main)			0.461***
			(0.147)
constant	3.779*	1.990	4.155***
	(2.174)	(1.614)	(1.559)
R2_Overall	0.947	0.961	0.955
N_Sample_Size	19.000	19.000	19.000
T_Sample_Size	11.00	11.00	11.00
Observations	187.00	187.00	187.00

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

TWD Models PR24 Econometric Consultation with Property Density

	TWD_Excl b/se	TWD6 b/se	TWD3 b/se
Ln(Mains)	1.055***	1.045***	1.072***
	(0.040)	(0.022)	(0.042)
Ln(Prty_Dsty)	-18.139***	-16.623***	-14.921***
	(2.261)	(1.562)	(2.259)
Ln(Prty_Dsty)^2	2.229***	2.055***	1.898***
	(0.263)	(0.180)	(0.264)
Ln(APH_TWD)		0.357***	
		(0.071)	
Ln(BoosStn_per_Main)			0.488***
			(0.148)
constant	30.876***	26.125***	25.065***
	(4.788)	(3.415)	(4.642)
R2_Overall	0.951	0.966	0.958
N_Sample_Size	19.000	19.000	19.000
T_Sample_Size	11.00	11.00	11.00
Observations	187.00	187.00	187.00

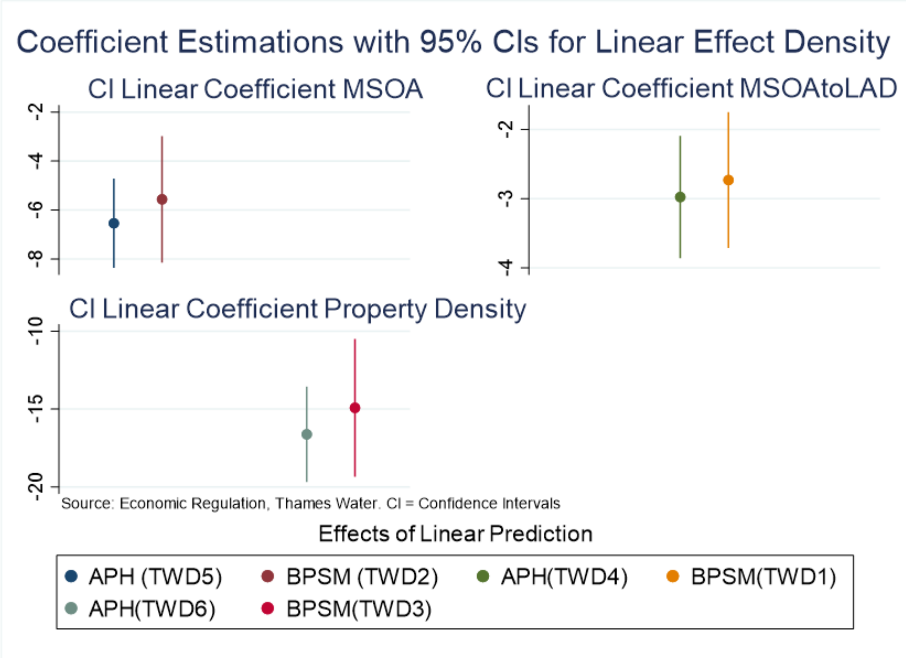
Source: Economic Regulation, Thames Water.
* p<0.10, ** p<0.05, *** p<0.01

This result provides further evidence of the collinearity issues that BPSM adds in the models, causing an overestimation of economies of density and therefore its impact on efficiency scores.

Finally, the overall statistical performance of the TWD models using APH [R^2 ranges between 0.961 to 0.966] is superior to the models with BPSM [R^2 ranges between 0.952 to 0.958] and the standard errors of the cost drivers in the model with APH are consistently minimised in any of the TWD models relative to the BPSM models, providing more confidence and precision in the estimated parameters with the models that use

APH (see standard errors in brackets in Table 5, or figure 7 that illustrates the confidence intervals for the linear effect of density across all the models[12]).

Figure 7



4. Overestimation of Elasticity of Density when using BPSM

The effect or implication that the estimated density coefficients has on the elasticity of density seems to be overestimated when BPSM is used in the TWD models (also in WW models). To explore this, we calculate the elasticity of density derived from the TWD models presented in the consultation and compare its results against the elasticity of density derived from the model that excludes either of the two drivers APH or BPSM, as presented in table 5 (for example, see models TWD_Excl, TWD2 and TWD5 in the top-left of table 5).

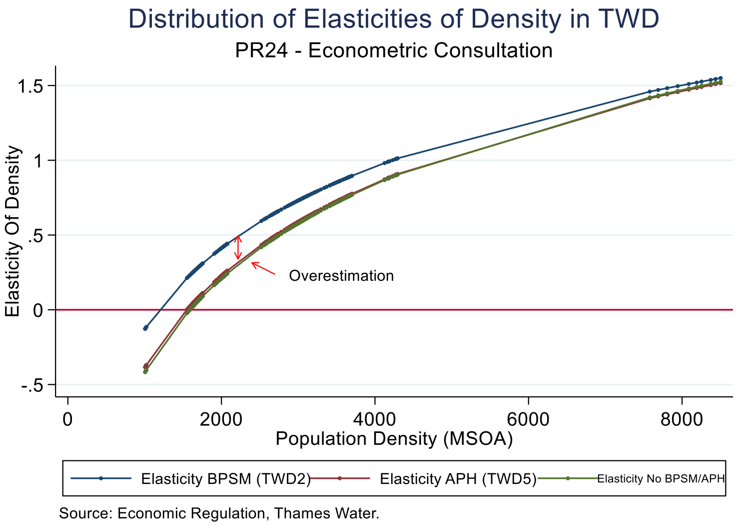
To calculate the elasticity of density for each company per year (ϵ_{D_it}) we use the expression below derived from the TWD models (TWD_Excl, TWD2 and TWD5):

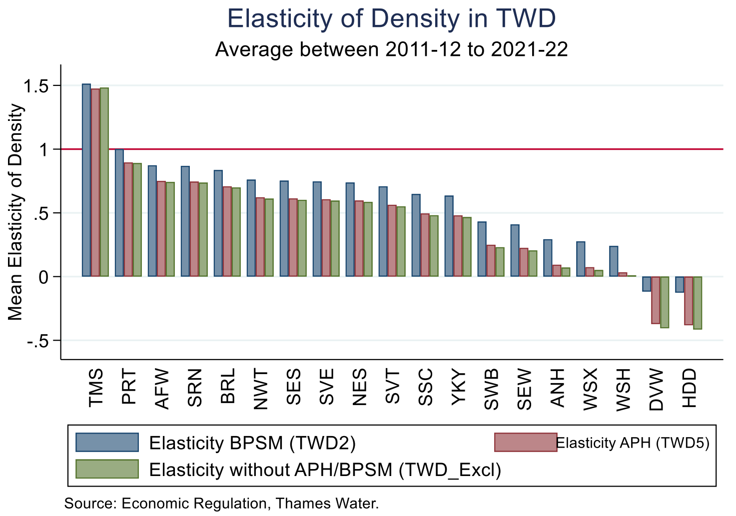
$$\epsilon_{D\_it} = \frac{\partial Ln(Botex_{it})}{\partial Ln(Density\_MSOA_{it})} = \beta_{MSOA} + 2\beta_{MSOA\_SQ} Ln(Density\_MSOA_{it})$$

[12] Figure 7 represents each of the different linear effects of the density drivers considered in the TWD models. Confidence Intervals (CIs) are depicted for each these coefficients. Larger CIs such as reflects a large variance of the estimated coefficient, which display a less precise estimator/effect/impact on botex. Among all the models in TWD, the ones with MSOA provides the lower Confidence intervals, and between these two models (TWD5 and TWD2), the model with APH provides the most precise estimator for density.

Figure 8 illustrates the results of the elasticity of density across companies using different model specifications TWD2, TWD5 and TWD_Excl. For example, the chart on the left depicts the gap between the TWD2 that uses BPSM (blue line) versus the models that include APH (TWD5, red line) and exclude both drivers (TWD_Exc, green line). The gap shows an indicative overestimation of the elasticity of density when the models use BPSM, particularly for companies with lower levels of density.

Figure 8





We can notice that the elasticity of density is quite consistent with the model that uses APH and with the one that excludes either APH or BPSM, suggesting the consistency of the effect that density has on botex costs when APH is used.

Similarly, the right chart on figure 8, shows the average using the 11 years of the sample for these elasticities across companies, showing that the elasticity for model TWD2 is always over the TWD5 and TWD_Excl models. We found that the same pattern shown in the charts in figure 8 applies to all the TWD models presented in the consultation as well as the all the WW models.

The average elasticity summary for the three cases explored in models TWD2, TWD5 and TWD_Excl at the industry level is presented in table 6 below as the quantification of the overestimation presented by model TWD2 that includes BPSM in figure 8 (in all models in TWD and WW, the industry average when using BPSM is always significantly above to the cases when APH or no APH/BPSM are used; the results are quite robust and consistent):

Table 6

Elasticity of Density	Industry Mean
TWD2 - BPSM	0.640
TWD5 - APH	0.486
TWD_Excl - No APH/BPSM	0.472

Sources: Economic Regulation, Thames Water

The effect that the estimated coefficient in density has on the industry efficiency scores when the models use BPSM or APH is significant. This helps to explain the significant swaps in efficiency rankings for some companies that are being affected by the inclusion of BPSM, whereas some others would get a fictitious efficiency gain.

We think that this evidence should be considered carefully as BPSM is not intended to capture what is supposed to capture regarding topography or energy intensity and on the contrary BPSM is correlated with population density. This “double” counting between BPSM and population Density, due to the lack of precision on what is BPSM actually capturing, and how the model struggles to isolate the real effect of density, seems to benefit some companies at the cost of others.

We think that BPSM is linked to the stock of capital of the company (e.g., asset intensity that the network or costumers demand as suggested in CEPA’s[13] report) and that the number of stations per Km of main is influenced by population density. In addition, it is not the number of stations per Km of main that is relevant to explain costs, rather it is the characteristics of those stations operating in the network, for example its capacity, that will determine its link with power costs[14].

We also think that BPSM role in the models should be questioned. Not only from the econometric, engineering, and economic point of view as previously discussed, but also from the regulatory incentives that the driver might be generating in the context of the desire for zero carbon emissions. We question the incentive that the BPSM variable has in the models regarding the efficiency use of energy. The BPSM variable may provide a perverse incentive for the use of energy consumption. We think that this perverse incentive is higher than APH.

We explore in more detail in the next section the effect that the number of booster pumping stations (stock) has in the models due to its high correlation with the scale driver (mains or properties).

[13] See CEPA, 'PR24 Wholesale Base Cost Modelling', March 2023, p. 60.

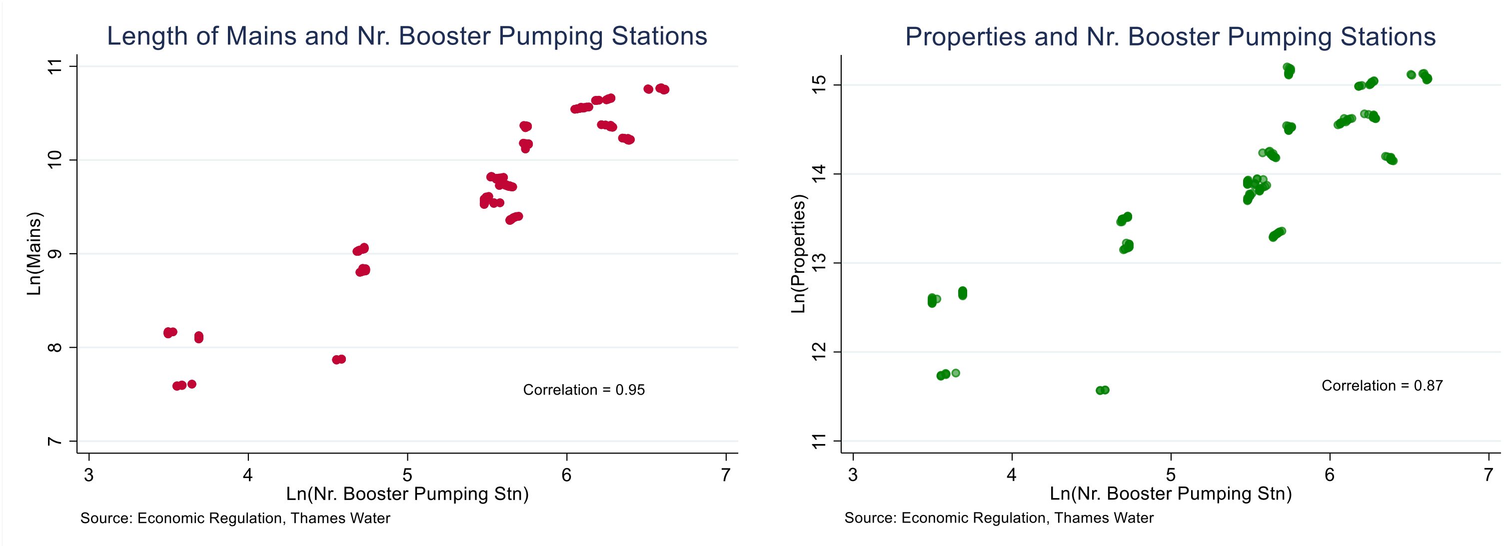
[14] Pumping stations are assets with an average asset life of 20 years. The structure of costs of a Booster pumping station is mainly driven by power costs. There is little capital maintenance on a year base due to the asset life of the stations. Most of the costs are related to power and operational side.

5. Correlation between the Scale Driver and the Number of Booster Pumping Stations

We include this section another view on how the Number of booster pumping stations, normalised by the length of mains (BPSM), impact the models. This will help us to understand further the effect that BPSM has in the models.

Figure 9 shows that the correlation between the Number of Booster Pumping Stations (stock) and the length of mains is high, 0.95. Similarly, a correlation of 0.87 is found with properties.

Figure 9



Given this high correlation between the total number of stations and Length of Mains (or properties) it could be argued that the stock of Number of Booster Stations (NST)[15] is a valid representation of the dimension or scale of the network as well as the Length of Mains or Properties drivers used in the TWD and WW models.

Based on these correlations and assuming that NPS is a good representation of a scale driver as indicated by the correlations mentioned before, we explore the effect that NST has in a hypothetical set of models for TWD and WW. We should expect a similar outcome in terms of the magnitude and sign of the coefficient when NPS is used instead of Mains or Properties.

[15] We use NST to refer to Number of Booster Pumping Stations and to distinguish It from the Number of Booster pumping Stations per length of mains (BPSM).

Table 7

TWD Models PR24 Econometric Consultation changing the Scale Driver

	TWD2 b/se	TWD_Nr b/se	TWD_Nr_M b/se	TWD_Nr_APH b/se	TWD5 b/se
Ln(Mains)	1.026***		0.593***		1.017***
	(0.050)		(0.134)		(0.019)
Ln(BoosStn_per_Main)	0.433***	-0.593***			
	(0.128)	(0.134)			
Ln(MSOA_Density)	-5.561***	-5.561***	-5.561***	-4.036**	-6.539***
	(1.316)	(1.316)	(1.316)	(1.603)	(0.928)
Ln(MSOA_Dsty)^2	0.393***	0.393***	0.393***	0.315***	0.445***
	(0.080)	(0.080)	(0.080)	(0.099)	(0.058)
Ln(Nr_Booster_Stn)		1.026***	0.433***	0.973***	
		(0.050)	(0.128)	(0.050)	
Ln(APH_TWD)				0.285**	0.411***
				(0.114)	(0.051)
constant	15.638***	15.638***	15.638***	10.022	16.573***
	(5.030)	(5.030)	(5.030)	(6.297)	(3.670)
R2_Overall	0.952	0.952	0.952	0.941	0.965
N_Sample_Size	19.000	19.000	19.000	19.000	19.000
T_Sample_Size	11.00	11.00	11.00	11.00	11.00
Observations	187.00	187.00	187.00	187.00	187.00

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

Table 7, presents the results when NPS is used in the models as a scale driver. We compare its results with the TWD2 and TWD5 models to see the implications and what can intuitively be learnt from this exercise.

The first column in Table 7 represents the TWD2 model of the consultation, whereas model TWD_Nr refers to the same TWD2 model but instead of using Mains as the scale driver it uses NPS (or Ln(Nr_Booster_Stn) in table 5). As expected, the coefficient of the scale driver remains the same at 1.026 as expected as well as all the other drivers included in the model, except for the coefficient of Booster Pumping Stations per Mains which is supposed to capture topography or energy intensity consumption.

The correlation matrix between the drivers in the model TWD_Nr are shown in table 8, suggesting that NPS does not have any correlation with BPSM and MSOA density:

Table 8

	Ln(NPS)	Ln(BPSM)	Ln(MSOA)
Ln(NPS)	1		
Ln(BPSM)	0.19	1	
Ln(MSOA)	-0.01	-0.60	1

In addition, the mean VIF derived from the OLS regression of model TWD_Nr without its square term is 1.44, which could lead to the false conclusion of non-collinearity issues. However, if we use the Conditional Number Index, there is evidence of mucollinearity issues between Density and BPSM[16] which can be intuitively derived from the correlation matrix (correlation of -0.60 in table 8).

The swap in the coefficient sign from positive to negative is again another way to see the concerning issues that BPMS brings in the models (see models TWD_Nr in table 7 and model WW_Nr in table 9, for example).

In the same table 9, we can see in the last two columns how the effect of the APH coefficient is quite robust and stable regardless of the scale driver used in the models, either with length of mains (Model TWD5) or number of booster pumping stations (Model TWD_Nr_APH), which corroborates the strength of the APH driver has from the econometric point of view.

Table 9

WW Models PR24 Econometric Consultation changing the Scale Driver

	WW3 b/se	WW_Nr b/se	WW_Nr_M b/se	WW_Nr_APH b/se	WW9 b/se
Ln(Properties)	1.052***		0.602***		1.041***
	(0.064)		(0.081)		(0.031)
Complexity36	0.003**	0.003***	0.003***	0.003***	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(BoosStn_per_Main)	0.509***	-0.717***			
	(0.169)	(0.095)			
Ln(MSOA_Density)	-4.684***	-4.443***	-4.486***	-3.246	-6.145***
	(1.404)	(0.923)	(1.190)	(2.211)	(1.238)
Ln(MSOA_Dsty)^2	0.301***	0.313***	0.300***	0.262*	0.384***
	(0.085)	(0.057)	(0.073)	(0.134)	(0.076)
Ln(Nr_Booster_Stn)		1.008***	0.432***	0.953***	
		(0.049)	(0.099)	(0.067)	
Ln(APH_TWD)				0.291**	0.359***
				(0.144)	(0.115)
constant	10.300*	11.655***	10.548**	7.504	13.173**
	(5.389)	(3.678)	(4.595)	(8.926)	(5.124)
R2_Overall	0.963	0.968	0.967	0.938	0.961
N_Sample_Size	19.000	19.000	19.000	19.000	19.000
T_Sample_Size	11.00	11.00	11.00	11.00	11.00
Observations	187.00	187.00	187.00	187.00	187.00

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

We find similar results for the Wholesale Water (WW) models. Table 9 illustrates the example using models WW3 and WW9 and the same conclusion on the robustness of the APH model against BPMS is found.

[16] If the Conditional number is greater than 20 or 30, as suggested by Greene (2012) or Belsley (1980), respectively, collinearity issues are suggested to be the case. In this case the condition number is 53.08, which is high. Moreover, the variance-decomposition proportions suggest that the sources of multicollinearity problems are between BPMS and Density, where the variance-decomposition of Density is quite high, 97%.

Final thoughts between APH and BPSM

It is widely acknowledged that the APH is a more appropriate cost driver than the BPSM to capture energy requirements to distribute water, due to topography conditions that companies face. Ultimately, water companies need to provide water in the most efficient way wherever the demand or customers are located.

The case for the BPSM as a cost driver is not underpinned by a clear engineering rationale. The BPSM is defined in general terms and consequently conceals the underpinning factors that are likely to be stronger drivers of cost – in particular the company-specific conditions, energy intensity, or the operational characteristics of the stations (e.g., topography).

APH is likely to be a clearer cost driver from an operational, engineering, and economic perspective.

The BPSM is not a robust driver from the engineering, economic, and econometric point of view as there is the potential that a single large pumping station with large capacity to pump water (hence higher levels of APH) could support or cover more efficiently a high-density area (e.g., to pump water in buildings and dense neighbourhoods etc.) or a spread area, exploiting their economies of scale in terms of capacity. Or it could be more efficient for a company to have more low-capacity pumping stations per length of main operating in areas with lower levels of density.

In other words, the current proposed models that use BPSM as a driver that intend to capture “topography” or energy intensity can be misleading as a proxy to explain power costs, as this is not describing the real characteristics of the Pumping Stations and how much on average they can reach/cover according to the area where they operate.

It is not about the number of Booster Pumping Stations per main; it is about the most efficient way to utilise the capacity of pumping stations and its design given companies’ specific circumstances to deliver and satisfy the water demand. This process will produce different levels of APH across the industry reflecting these efforts to pump water regardless the location of customers. This is what a driver intending to capture distribution topography or energy intensity use, should reflect.

If an alternative or substitute for BPSM is used we believe that Capacity is a more robust and engineering driver that should be used alongside APH, as we proposed in our January econometric model’s submission. However, for clarity and consistency with the engineering rationale we think that APH is still the most objective driver to represent energy intensity requirements.

Q3.3) Do you agree with our approach to modelling population density?

We strongly agree with the approach of modelling the relationship between cost and density as a non-linear relationship (i.e., using a linear and a quadratic term). This approach has a sound economic and engineering rationale as well as strong econometric evidence, which we set out below. The non-linear relationship between cost and density has been successfully used at PR19 and was accepted by companies, Ofwat and the CMA.

The non-linear relationship between cost and density reflects the following: at low density areas, additional density is expected to reduce costs due to economies of scale (i.e., lower unit cost) in production, and other factors such as proximity to the supply chain. As density increases, the marginal benefit of these factors diminishes, and, in addition, new costs come into play. These costs are due to factors such as:

- Higher maintenance costs of below ground assets due to hard surfaces and higher standard of roads to reinstate.
- Congestion of underground assets complicates access.
- More leaks due to pressure issues. Harder to locate leaks due to more intricate network and hard surfaces.
- A higher proportion of 'critical' assets.
- Lower productivity due to slow traffic, distance to refuse sites and restricted storage space.
- Local authority charges.
- Higher customer management costs as more customers affected by incidents.

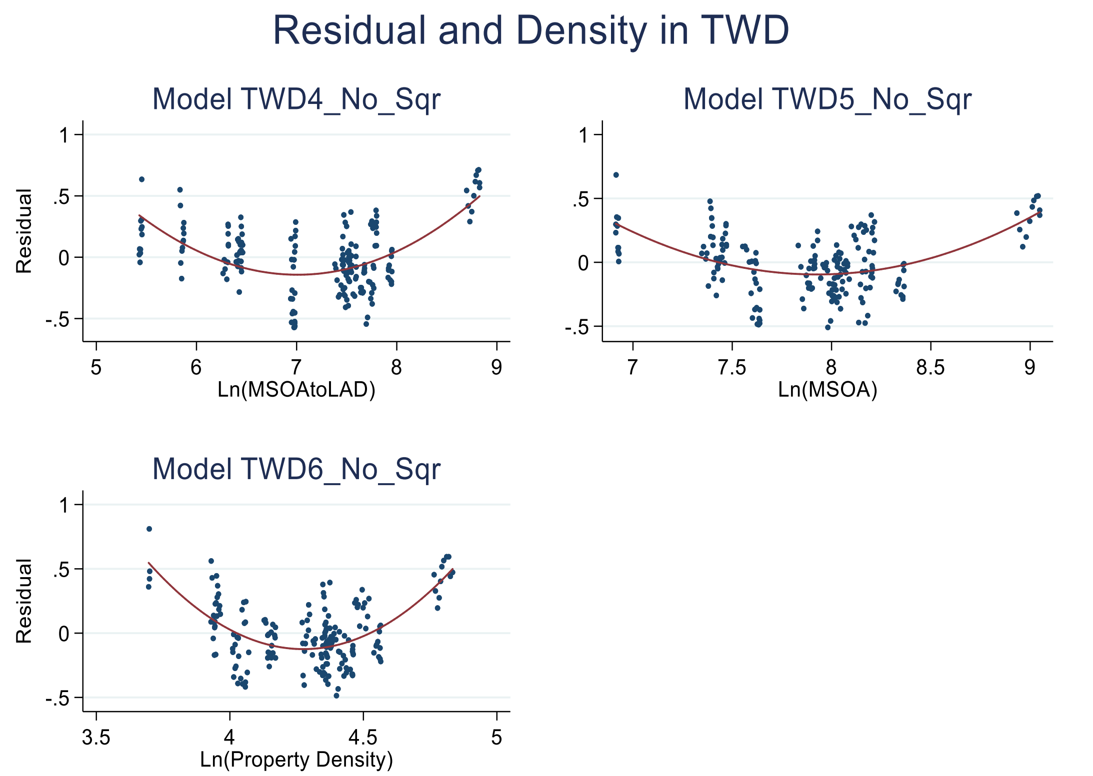
In addition to the economic/engineering rationale set out above, there is also clear econometric evidence to support such relationship: the square term of density is statistically significant in all models, and it improves their fit.

We have undertaken a residual analysis to further demonstrate the importance of the squared term in the models. In a residual analysis we examine the relationship between the model's residuals and its explanatory variables. For a well specified model we expect to find no relationship between the explanatory variables and the model's residual. That is, the residuals must have an expected value of zero conditional on the explanatory variables.

Figure 10 provides a residual analysis. It shows the relationship between the level of density and models' residuals in models where the squared term is excluded. The figure shows that there is a U-shape relationship between the level of density and the models' residual, counter to what we would expect to find in a well specified model.

The figure shows that the residuals are not independent of the level of density. At low and high levels of density the residuals have a positive expected value whereas in the middle they have a negative expected value.

Figure 10



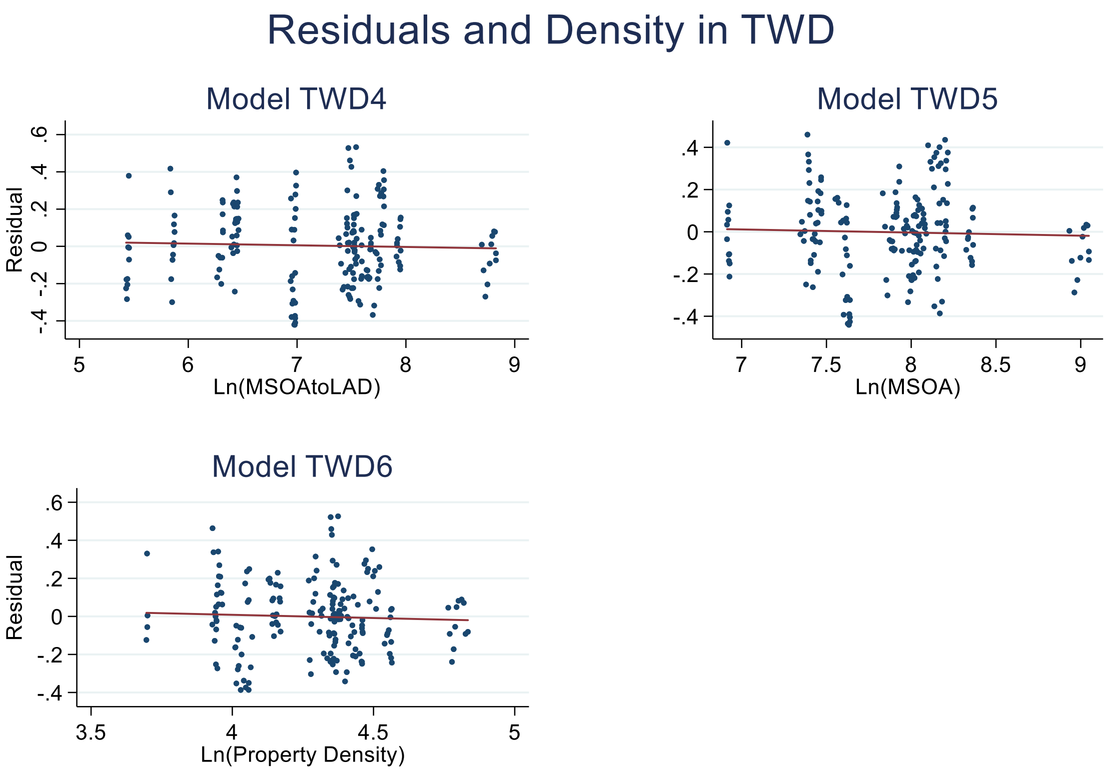
Source: Economic Regulation, Thames Water. Note: The No_Sqr, means model do not include Square Term of Density.

Introducing the square term of density in the models as proposed in the consultation, the residual analysis reveals that the U-shape pattern for any type of density disappears. We show this in figure 11.

This evidence suggests that the residuals are independent on the level of density either property or population once the square term of density is introduced, indicating that the issue is resolved, and models significantly improved[17].

[17] We did the same analysis for WRP and WW models and the same pattern of a U-shape relationship is found as it is described for the TWD example presented in this response in figure 10.

Figure 11



Source: Economic Regulation, Thames Water.

Given the engineering and economic rationale for a non-linear relationship between cost and density, and the strong statistical evidence for a squared density term in the model (as evidenced through its statistical significance, its positive impact on overall model quality and through the residual analysis), we consider that it is important to continue using models with a squared term of density.

Which of the three proposed population density variables do you support?

a. Weighted average density – LAD from MSOA

b. Weighted average density - MSOA

c. Properties per length of mains

We support the weighted average density – MSOA measure. Given its advantages over the other measures we consider that it would be appropriate to use it as the only measure of density in the models. We expand below.

Why we consider that the weighted average density - MSOA measure is superior to the properties per length of mains measure.

There are weaknesses to the 'properties per length of mains' measure, which we consider make it clearly inferior to the other measures, and unfit for purpose. As acknowledged in the consultation, the properties per length of mains measure is not fully exogenous, while the other two measures are. But crucially, this measure does not capture the presence of sparse and dense sub-areas within a company's supply area. It is important to capture the presence of sparse and dense sub-areas. They determine the economies of scale the company is faced with. Water companies serve large geographical areas. The simple average density across the entire supply area does not capture economies of scale and therefore is not a useful measure of density for our purposes. Two water companies with the same overall density can have very different opportunities for economies of scale if one has sparse and dense region and the other has 'moderately populated' sub-areas throughout. The weighted average density measures would suitably differentiate between these two regions. The properties per length measure would not.

To summarise, properties per length of mains does not capture the effects we are interested in, namely, concentration of demand/connections that allow operation of large assets, hence is unfit for purpose and should be discarded, particularly given that there are superior alternative measures.

Why we consider that the weighted average density - MSOA measure is superior to the weighted average density – LAD from MSOA measure[18].

Ahead of PR19, Ofwat and the sector worked together to improve the PR14 measure of density, property per length of mains. Various measures were considered, such as concentration measures (e.g., HHI, Gini coefficient) and urbanity/rurality measures. The LAD-based weighted average density was considered the most appropriate measure of density (as it directly captures local population centres that can provide economies of scale in supply); it was also found to be the most statistically robust amongst all measures considered.

We consider that the weighted average density – MSOA measure provides a further improvement. Using MSOAs has two advantaged over LADs. One advantage, as noted by Ofwat and others, is that the boundaries of MSOAs change less often than the boundaries of LADs. This ensures that at each point in time we have a more accurate geographical mapping of population to water companies, a lower risk of errors due to changing boundaries and greater consistency of data and density measure over time.

The other advantage lies in the enhanced granularity of the data. As stated by CEPA referring to the MSOA variable: "this uses more granular data that may provide a more accurate picture of the relative density between company areas and may be less sensitive to changes in the dataset over time".

[18] For a more illustrative explanation of the advantage of using MSOA over LAD, please see our answer to the equivalent question in the network plus question Q4.3, where some maps are shown.

Given these advantages of the MSOA, and the conceptual proximity of the two weighted average density measures, we consider that choosing the one that is based on MSOA boundaries as the only density measure in the model is appropriate.

Q3.4) Do you agree we should collect additional data on the number of reservoirs that are designed as high-risk by the Environment Agency and Natural Resources Wales?

Yes, we agree. We think is sensible to collect this information.

Do you have a view on the appropriateness of capturing a variable for reservoir inspection and maintenance requirements under the Reservoir Act 1975 in the water resources plus models?

We think that a similar approach of what we proposed in our January submission should be taken as an input. For example, taking the capacity of the reservoirs that are under the Reservoir Act 1975.

Wholesale Wastewater Network Plus Questions

Q4.1) Do you agree with our proposed set of wastewater network plus base cost models?

We welcome Ofwat's inclusion of the rainfall rate as well as the use of Load as the scale factor in WWWNP. These are variables we have been suggesting since the 2021 December Base Cost Consultation.

We are concerned with the treatment of density in SWC models. We think that the measure of density must capture the significant heterogeneities of density within and between companies for accurate cost benchmarking. It is also crucial to capture the non-linear relationship between density and costs, as in water, otherwise efficiency assessment and companies' funding would be distorted.

Regarding the proposed drivers for economies of scale in SWT models and WWWNP, we think that the percentage of load treated at size bands 1 to 3 and % Load treated in STWs >100k should be used. This is in line with the PR19 and CMA models.

We think that the driver WATS is a promising driver to reflect the industry structure of STWs, but we are concerned about the consistency within the calculation of the driver that uses two methods. One method for bands 1 to 5 that uses a weighted simple average (WSA) and the second method that uses each large sewage treatment works or weighted individual contribution (WIC). We provide more detail in our response to question Q4.2

We welcome the inclusion of *Load treated with ammonia consent ≤ 3mg/l* as the variable that captures the sewage treatment complexity across the industry. All companies face NH3 restrictions consents, and the driver has a strong effect in base costs in SWT and WWWNP models.

Finally, regarding the inclusion of the proportion of coastal areas, we think that this is best addressed via a Cost Adjustment Claim. Some models are quite sensitive to the inclusion of the driver. After some sensitivity tests, some models swap the sign of the coefficient, which is a concerning issue regarding the stability of the driver.

For these reasons, we neither agree nor disagree with the current suite of proposed models for PR24.

Q4.2) Do you agree with our approach to modelling economies of scale at sewage treatment works?

Which of the three proposed explanatory variables do you support?

a. Percentage of load treated in STWs bands 1 to 3

b. Percentage of load treated in STWs serving more than 100,000 people

c. Weighted average sewage treatment works size

We agree with the use of the first two measures: *Percentage of load treated in STWs bands 1 to 3* and *Percentage of load treated in STWs serving more than 100,000 people*.

We see these two drivers as a continuation of the models used at PR19 by Ofwat and the CMA. Although, the *Percentage of load treated in STWs bands 1 to 3* is not statistically significant in the SWT models, it is significant in the WWWNP models. This provides evidence of the relevance of the driver across the wastewater network.

We have concerns with respect of the *Weighted Average Sewage Treatment Works Size* (WATS). On the one hand, it uses a fuller set of information than the measures above; it reflects the weighted average size of STWs for each company, using all information available in companies' Annual Performance Reports. As such, we consider that it has merit and can complement the two other measures proposed. Moreover, the driver is statistically significant, stable, has the correct sign and magnitude, and improves the R^2 of SWT models.

However, there is an inconsistency in measuring the variable. The variable is a combination of two methods: one for bands 1 to 5 (simple average of treatment works) and a second method for works above band 5, that captures the contribution of each large STW individually.

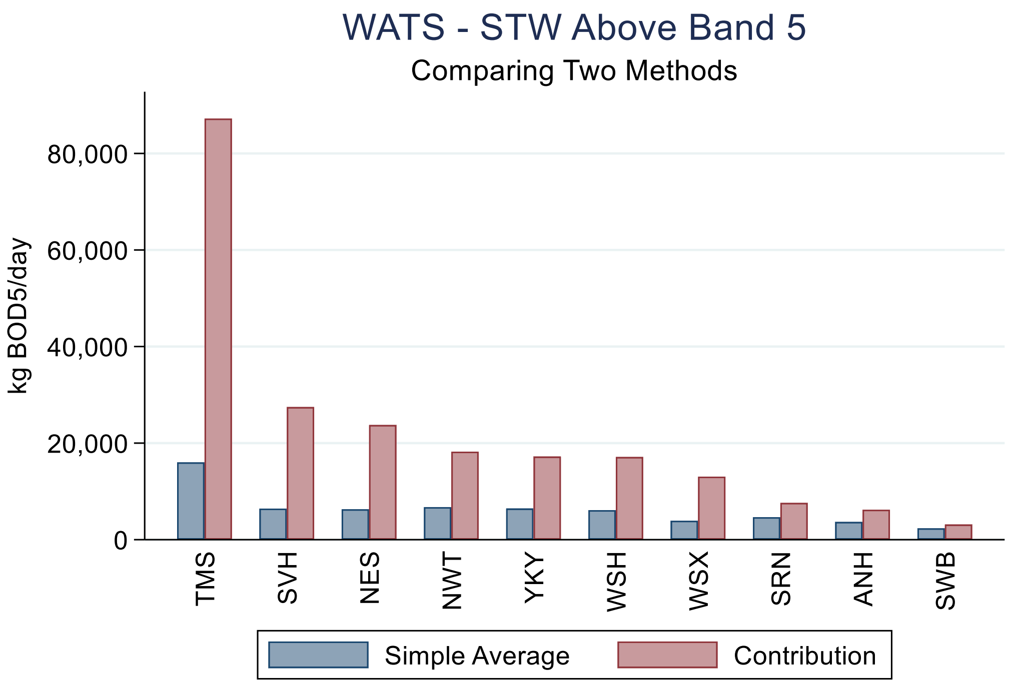
CEPA also recognises this - "*a limitation of this variable is the lack of a complete dataset for all STWs which means that the calculation of the WATS variable involves two distinct approaches for STWs in bands 1-5, and STWs in bands above 5*[19]".

We investigate the WATS variable to check for the potential inconsistency across bands and companies. We follow the same definition of WATS in Appendix 3 of the Base Econometric Consultation (p. 74) and applied the same method (simple average) used in *band 1-5 to sewage works above band size 5* and compared it with the method (contribution) used in the variable WATS, that captures the contribution of *each* large sewage treatment work. Figure 1, illustrates the differences between

[19] See CEPA, PR24 Wholesale Base Cost Modelling, p. 85.

the two methods and the mathematical expressions below show in more detail the methods:

Figure 1



Source: Economic Regulation, Thames Water

$$Weighted\ Individual\ Contribution\ (WIC) = \sum\nolimits_{ijt} (Load\ recevied\ by\ STW_{ijt}\ \ above\ size\ band\ 5) * \left(\frac{Load\ Received\ by\ STW_{ijt}}{Total\ Load_{it}} (\%)\right)$$

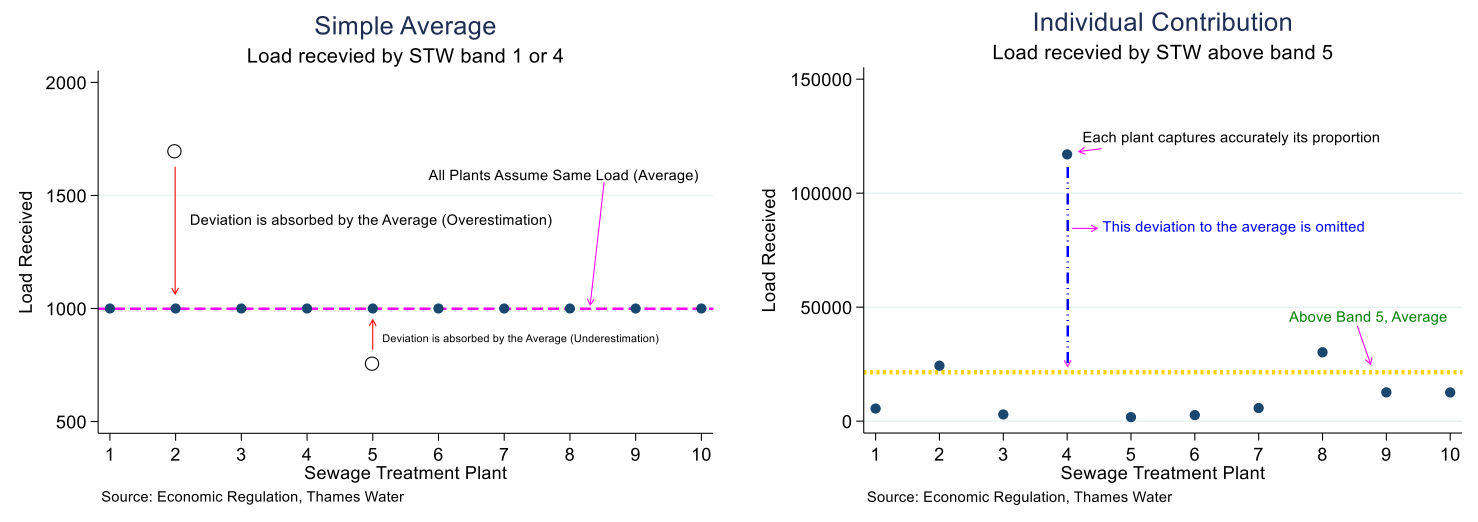
$$Weighted\ Simple\ Average\ (WSA) = \left(\frac{Load\ Received\ by\ STWs\ in\ size\ above\ band\ 5_{it}}{Number\ of\ STW\ above\ size\ band\ 5_{it}}\right) * \left(\frac{Load\ Received\ by\ STWs\ in\ size\ above\ band\ 5_{it}}{Total\ Load_{it}} (\%)\right)$$

Figure 1 illustrates the potential impact of the inconsistency between using the WSA and WIC within the same WATS calculation. If we extrapolate this to bands 1 to 5, the impact on the WATS could be significant. We do not have this evidence at the moment, but this illustration can inform us about the potential underestimation of the WSA versus the WIC. This inconsistency in calculating the driver could bring material implications for companies. Moreover, we consider that the WATS, by its construction, may be over-stating the benefits of economies of scale at large STWs. The formula assumes a linear relationship between plant capacity and costs. But economies of scale are generally not linear: the marginal benefits of scale tends to be high at small scale and low at high scale. Figure 1 provides some confirmation of this, as it shows a very sharp increase in the value of the WATS for companies with large STWs. One option to mitigate this would be to assume a log relationship between plant scale and cost by using log of the capacity within the summation formula. This would be more accurate than the use of log after the summation.

Coming back to the inconsistency issue, we illustrate in a different way our concern between the two methods used in the WATS. Figure 2 depicts an illustrative example of the simple average and individual contribution methods used in WATS. In the simple average method, in the left chart of figure 2, we can see that each plant assumes the same simple average. However, in reality, it could be the case that the plant's position is higher or lower than the average, and the deviation from the average could be high as opposed to the individual contribution method used in the Load received above band 5, which omits this issue.

This deviation in the WATS could be considered as an inconsistency across the two methods used to calculate it and it could be absorbed in the model as a measurement error[20].

Figure 2

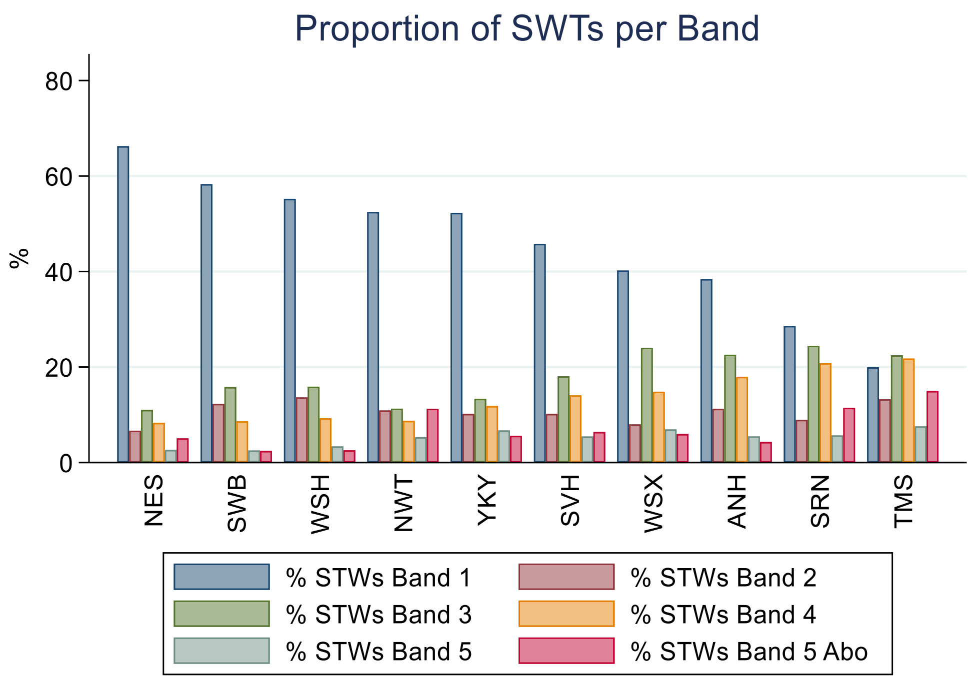


[20] Some of the sources of measurement error are usually coming from:

- Incorrect response to a survey question
- Incorrect coding of a correct response
- The use of a correctly measured variable as a proxy for another theoretically valid but unobserved variable (See Microeconometrics Cameron and Trivedi (2005))
- Wooldridge (2010), *Econometric Analysis of Cross Section and Panel Data, mentions that measurement error is an issue only when the variables for which the econometrician can collect data differ from the variables that influence decisions by individuals, firms and so on*. For example, the contribution of each STW in each Band has a different weight or size on the total specific Band Load received (e.g., Band 1). This contribution or weight of STW is determined by different factors such as the size of the plant, its location, population density among others that influence companies' decisions in establishing a STW in a particular area. The individual contribution of each STW for all bands 1 to 5 is information we do not observe or collect in the current measure of WATS (we only have for large STW). Instead, we are using the average of STW for band 1 to 5. Companies do not take decisions on the average of a STW in a particular band, it depends on the contribution, location, density, and other factors. Hence, the average used for STW in bands 1 to 5, differs from the variables or factors that really affect the economic behaviour of companies in installing a particular STW and hence its real contribution to the Band. This is a source of measurement error.

This could cause an over or underestimation for the results between bands 1 to 5. There is a significant difference in the distribution of the number of STWs across the industry (see figure 3). For instance, NES has nearly 70% of their number of STWs in band 1, whereas SRN or TMS have only around 20%. In addition, the structure of this distribution is significantly different. By using the simple average method, some of the companies with large amount of STWs in lower bands could get a hypothetical benefit in efficiency by using the simple average across bands 1 and 5, as the deviation could be minimised, whereas companies with a small proportion of STWs could be affected by not having a consistent measure across the lower bands.

Figure 3



Source: Economic Regulation, Thames Water

In the econometric models, this inconsistency or deviations observed in the simple average method could be seen as a source of measurement error. To illustrate its potential effect in the models, suppose the following simple population relationship:

$$Cost = \beta(WATS_{Consistent}) + \epsilon$$

But we only have data on:

$$\widetilde{WATS_{inconsistent}} = (WATS_{Band\ 1-5} + Deviation) + WATS_{Above\ Band\ 5}$$

If we consider this deviation (or inconsistency with respect to Load above band 5) as an error of measurement (u) and substituting this in the previous equation, we have:

$$Cost = \beta \widetilde{WATS_{inconsistent}} + (\epsilon - \beta u)$$

If u is positively correlated with $\widetilde{WATS_{inconsistent}}$ then the estimator of β is biased and inconsistent.

We have tested the following economies of scale measures for the presence of measurement error:[21]

- Pct Band 1 to 3 (used at PR19 and being considered for PR24)
- Pct Band 6 (used at PR19)
- Pct STWs larger 100k (being considered for PR24)
- WATS (being considered for PR24)

The results are presented in table 1:

Table 1

Delgado and Manteiga Test for Measurement Error

Ho: No Measurement Error

Number of bootstrap sample: *5,000,000*

	Pct Band 1-3	Pct Band 6	Pct SWTs large 100k	WATS
P-value	0.098	0.502	0.017	0.007

Source: Economic Regulation, Thames Water

We find that for drivers PCT Band 1 to 3 and Band 6 (used at PR19) there is no statistical evidence of measurement error. However, for the two new drivers, PCT SWT larger 100k and WATS, the null is rejected, suggesting the presence of measurement errors. In particular, the WATS is fully rejected with a p-value 0.007 lower than 0.01. We run the bootstrap resampling up to 5,000,000 iterations to make the p-values quite robust for each of the drivers.

We suggest that the Percentage of load treated in STWs bands 1 to 3 and Percentage of load treated in STWs serving more than 100,000 people provide more clarity and confidence in how the effect of STWs has on costs. However, based on the test results Pct Band 6, is more robust than Percentage of load treated in STWs serving more than 100,000 people, this could be considered as an alternative.

[21] Our test is based on Delgado, M. A., and W. Gonzalez Manteiga (2001). Significance testing in nonparametric regression based on the bootstrap. Annals of Statistics. There is an implementation in Stata (see at the Stata Journal) where the test can be easily carried out. https://journals.sagepub.com/doi/pdf/10.1177/1536867X20931002

The two drivers nicely complement one another and capture the key impact of economies of scale, each capturing one.

Should Ofwat decide to use WATS, it must triangulate its result with other, more robust, measures and provide the appropriate weights given the potential implication of measurement error in models using WATS or Pct STWs larger 100k. Failure to do so risks a biased model with inappropriate cost allowances for companies.

Q4.3) Do you agree with our approach to modelling population density?

We strongly disagree with characterising the relationship between cost and density as linear in the Sewage Collection cost models. We provide an economic and econometric explanation in the answer to question Q4.4. of why this linear effect is not the objective approach to model the impact of density in the models proposed.

Which of the three proposed explanatory variables do you support?

d. Weighted average density – LAD from MSOA

e. Weighted average density - MSOA

f. Properties per sewer length

We support the **weighted average density** – **MSOA** measure. Given its advantages over the other measures, we consider that it would be appropriate to use it as the only measure of density in the waste models. This is consistent with our response to the same question in water.

Why we consider that the weighted average density - MSOA measure is superior to the properties per length of mains measure*.*

We consider that population density is more beneficial to understanding operating costs than property density, for the simple reason that it is the wastewater produced (load) by people that generates the cost to operate.

As acknowledged in the consultation, the properties per length of mains measure is not fully exogenous, while the other two measures are. But crucially, this measure does not capture the presence of sparse and dense sub-areas within a company's area.

Waste companies serve large geographical areas. So it is important to capture the presence of sparse and dense sub-areas as they determine the economies of scale the company is faced with.

Variances in population density versus property density per sq km take into account not only the property type i.e., 1 bedroom starter homes versus maisonettes, flats or large multi bedroom houses where the occupancy number will be greater but also a bit more about the demographics i.e., areas may vary in terms of occupancy based on possibly house or location value where it is more common for single occupancy of homes versus multiple occupancy due to individuals' personal circumstances.

For example, in wastewater hydraulic modelling the approach is always based on occupancy (population density) as it is not possible to derive wastewater usage profiles from just a house (property) count.

A property might not reflect the weight it has on the network. For example, when a connected property is counted as one, it usually means that on average this property has 3 or 4 people living in it. However, sometimes a property could be considered as one connected property, but in reality, there are more than 3 people living in the property given that it could be a building where several households or flats are occupied (e.g., student accommodations, a private complex of residential buildings, or commercial office buildings like the financial sector in London) or a house that has been converted into several flats. These properties could be counted as one connected property therefore missing the real weight or impact that the property has on the network.

This impact is better captured with the weighted average population density measure. The economic output produced by a sewage collection process is the volume of wastewater collected through the length of the sewage network, which is mainly driven by the load produced by people. A weighted population density driver captures the impact that each person adds on the collection process through the load produced.

Furthermore, population density captured by MSOA areas provides a significant disaggregation advantage against the LAD or property density variables as explained before in question Q3.3. A company might have a very limited number of LADs units when compared to other companies.

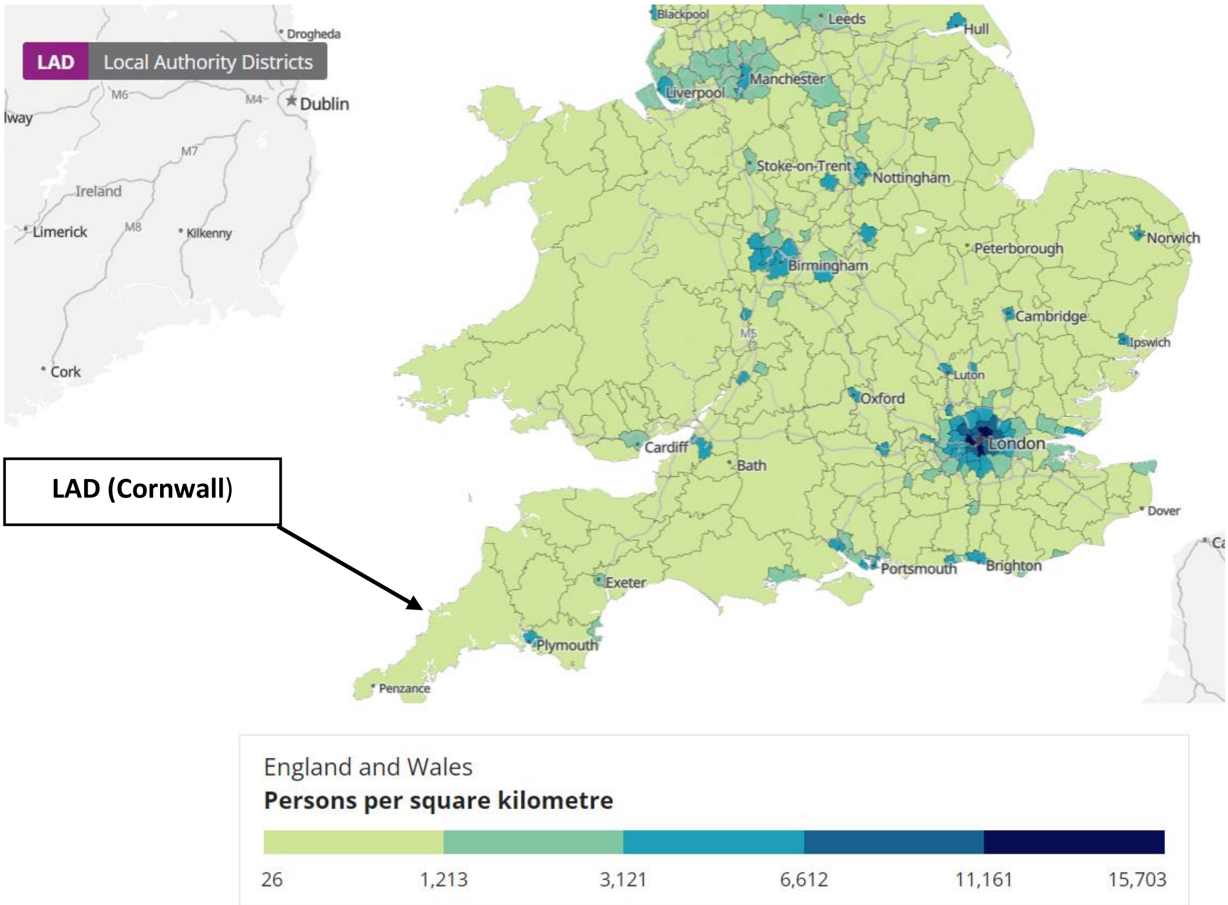
To summarise, the simple properties per length of mains measure does not properly capture concentrations of demand/connections that allow operation of large assets. It is inferior to the alternative measures and should be discarded.

Why we consider that the weighted average density - MSOA measure is superior to the weighted average density – LAD from MSOA measure.

With respect to the Population Density drivers, LAD and MSOA, we think that MSOA is a stronger measure. We consider that the weighted average density – MSOA measure provides a further improvement over the LAD based measure. Using MSOAs has two advantages over LADs. One advantage, as noted by Ofwat and others, is that the boundaries of MSOAs change less often than the boundaries of LADs. This ensures that at each point in time we have a more accurate geographical mapping of population to waste companies, a lower risk of errors due to changing boundaries and greater consistency of data and density measure over time.

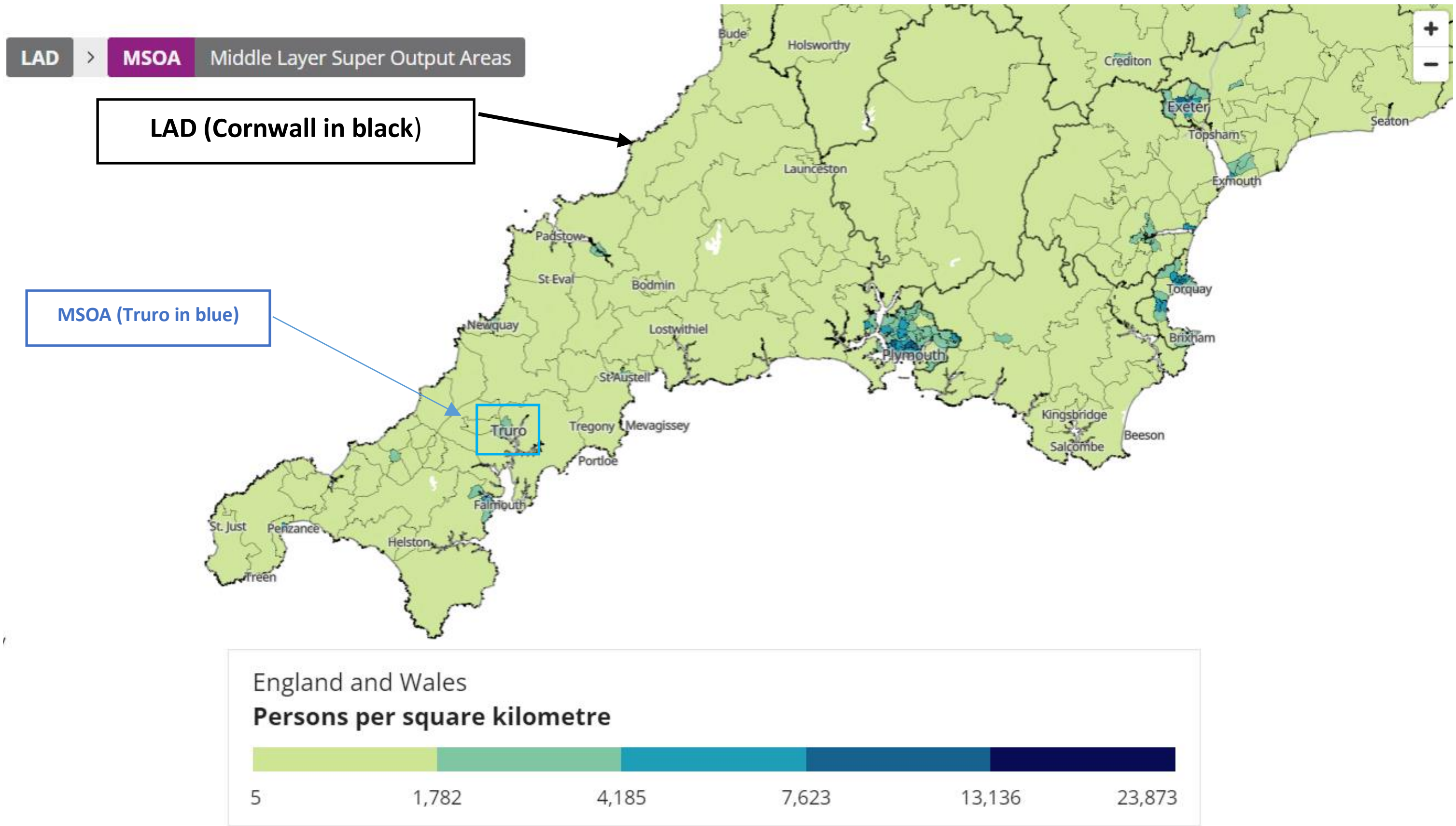
The other advantage lies in the enhanced granularity of the data. As stated by CEPA referring to MSOA variable: "*this uses more granular data that may provide a more accurate picture of the relative density between company areas and may be less sensitive to changes in the dataset over time*". To illustrate the MSOA advantage over LAD, the maps 1 and 2 below depict some of the benefits[22]:

Map 1



[22] The following maps are extracted as a screen shot from the interactive tool from ONS based on census 2021: https://www.ons.gov.uk/census/maps/choropleth/population/population-density/population-density/persons-per-square-kilometre More information on deprivation and other variables is also found in the link. The tool clearly shows the aggregation levels between LADs and MSOA areas.

Map 2



Map 1 illustrates all the LADs in England and Wales, which can provide a very granular and accurate picture if the aim is to analyse England and Wales as a country. If this is the case, the LAD provides a significant level of disaggregation to understand heterogeneities across the country.

However, if the aim is to analyse and compare regional areas, the LAD could be limited as some areas in England and Wales tend to have a smaller number of large LADs. In addition, at the LAD level opportunities for economies of scale arising from pockets of dense populations will be missed. For example, in Map 1 the density of the Cornwall-LAD is shown as uniformly low, whereas in map 2 pockets of denser populations areas like the case of MSOA-Truro can be seen.

Given these advantages of the MSOA, and the conceptual proximity of the two weighted average density measures, we consider that choosing the one that is based on MSOA boundaries as the only density measure in the model is appropriate.

Q4.4) Do you agree with our proposal to assume a linear relationship between population density and sewage collection base costs?

We strongly disagree. Operating a network in dense areas influences operating and maintenance costs. A sewage collection model must distinguish the different levels of densities that each company faces to assess objectively how an efficient company behaves and performs.

We show that there are economic/engineering and econometric reasons why is important to introduce a nonlinear effect of density in the SWC models. These reasons are:

- Economic and Engineering Rationale of a Nonlinear Density Effect in SWC
- Nonlinear Effect of Density in SWC: Econometric Results
- Elasticity of Density in Sewage Collection
- Functional Form Test: Restricted (Linear) versus Unrestricted (Nonlinear)

We now explain in more detail each of these reasons.

<u>Economic and Engineering Rationale of an operating a dense area in SWC</u>

The levels of density across the wastewater industry differ from company to company. Some of examples of why a denser area might tend to put more pressure on the wastewater sewage network reflecting more complicated operating environment than an area with low levels of density are:

- **Consequence of Failure**: Companies' might have a large proportion of its large sewerage assets in denser areas where the consequences of failure are high resulting in additional proactive preventive maintenance in expensive areas.
- **Size Asset Location**: In more dense areas there is likely to be a higher proportion of sewers crossing railways (over and underground) which will result in increased costs associated surveying and maintaining sewers crossing railways.
- **Type of Asset**: A company might have combined sewers in one part of its operating area (e.g., London), whereas elsewhere a different system can be operating (e.g., separate foul and surface systems). For example, a company where more than 50% of its systems is operated by a combined sewers might have 50% more flooding incidents per property than an area run by a different system, but also it might have a quarter of the number of blockages per property.

- **Cost of Work**: Failures in the network tend to be more expensive to fix in denser areas because it is harder getting access (e.g., cars parked over manholes, or not being able to access manholes around the backs of properties), and work is more complex and longer (e.g., it is harder digging up a street than a field).
- **Food Service Establishments and Commercial Properties**: Concentration of fast-food outlets will generally result in more blockages from fats and greases. In high density urban areas this type of commercial establishments are more agglomerated.
- **Transient Residency and Transport Infrastructure Networks:** In operating areas with large number of airports and other transport hubs (Eurostar, London Airports Network (5 airports), Birmingham, etc.) an impact due to incoming passengers to these areas prior to moving to their final destination impacts costs significantly. There is an incoming constant flow which is a continuous process with different peaks at the year (holidays, summer etc). An example of this was found during covid when passenger numbers dropped significantly in transport hubs.

Nonlinear Effect of Density in SWC: Econometric Results

We consider that there are strong reasons to include nonlinear effects in the sewage collection models.

We note that in the PR19 Water re-determinations the CMA included non-linear effects in one of their sewage collection models (SWC2). The CMA found that[23]:

- "*In sewage collection population density may have two opposing effects. These effects may vary according to the level of density. One way to capture these opposing effects is to include non-linear terms of population density. Therefore, we think it makes economic and engineering sense to include the squared term of population density*" and continues
- "*We found the SWC2 coefficients for both the weighted average of population density and its squared term to be statistically significant. The coefficients were also of the expected sign (the former was negative, and the latter was positive). Indeed, the population density terms in our SWC2 model suggested a similar effect to that of population density on wholesale water costs. At lower levels of density, scale economies are strong and therefore increasing density reduces costs. However, the positive effect of the quadratic term suggests that as density rises its*

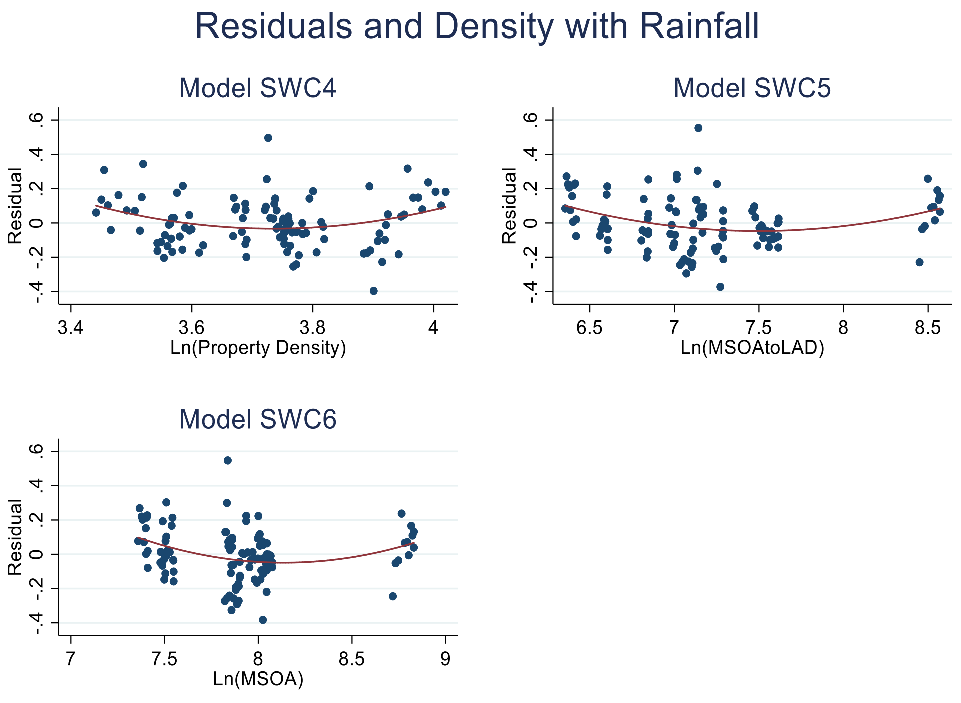
[23] CMA PR19, FD, p. 161, paragraph, 4.177 to 4.179.
https://assets.publishing.service.gov.uk/media/60702370e90e076f5589bb8f/Final_Report_---_web_version_-_CMA.pdf

negative impact on costs decreases, ultimately becoming positive at high values of density".

For these reasons the CMA decided to include the square term of weighted population density in SWC2 model. Moreover, the CEPA report to Ofwat also found that models with nonlinear effect of density were statistically significant supporting the evidence found by the CMA.

We have calculated the residuals of the SWC econometric models proposed in the consultation (without the square term of density) to understand its relationship with the corresponding measures of density. Figure 4 shows the relationship between the level of density and the residuals in models where the squared terms is excluded. The figure shows that there is a U-shape relationship between the level of density and the models' residual, counter to what we would expect to find in a well specified model (very similar results are obtained for models SWC1 to SWC3)[24].

Figure 4



Source: Economic Regulation, Thames Water.

With this initial evidence on the potential U-shape of residuals related to density, we move to the stage where the square term of density is introduced and tested in the econometric models to assess its real impact on botex costs.

---

[24] Excluding the most density company also suggest the U-shape relationship, particularly when the weighted population density is used. We can provide these results if requested.

Table 2 presents the econometric results using Ofwat's consultation format. We notice that across all the specifications the R^2 increases when compared to the proposed suit of models in the consultation that exclude the nonlinear (square) term of density.

The models that contain property density in table 2 (SWC1 and SWC4) show a lower level of statistical significance in general, when compared to the models that use population density and also these models suggest that there are some specification issues (see value of RESET test). This reinforces the preference to use population density and discard the use of property density.

Regarding the models that include the annual urban rainfall (SWC4, SWC5 and SWC6) we see that the driver adds valuable information into the models and improve the overall performance with higher levels of R^2 among the models proposed. The statistical level of significance of the cost drivers and most of the sensitivity tests tend to be more robust compared to models that exclude rainfall.

It is worth noting that in all models that use the nonlinear effect of density we find that at lower levels of density, density economies are strong, which is suggested by the negative sign of the linear density coefficient. The square positive effect of the quadratic term of density suggests that as density rises its negative impact on costs decreases, becoming positive with high levels of density (diseconomies of density), which is what the CMA finds at PR19.

Table 2

Square Term of Density in Sewage Collection Models

Cost driver	Explanatory variable	SWC1_SQ	SWC2_SQ	SWC3_SQ	SWC4_SQ	SWC5_SQ	SWC6_SQ
Scale	Sewer length (log)	0.794***	0.847***	0.852***	0.827***	0.857***	0.865***
		{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Topography	Pumping capacity per sewer length (log)	0.308**	0.594***	0.554***	0.290*	0.564***	0.526***
		{0.033}	{0.000}	{0.000}	{0.051}	{0.000}	{0.000}
Density	Properties per sewer length (log)	-8.635			-13.172*		
		{0.159}			{0.053}		
	Properties per sewer length (log) 2	1.302			1.904**		
		{0.115}			{0.037}		
	Weighted average density - LAD from MSOA (log)		-2.291**			-2.042***	
			{0.041}			{0.001}	
	Weighted average density - LAD from MSOA (log) 2		0.169**			0.154***	
			{0.021}			{0.000}	
	Weighted average density – MSOA (log)			-5.051*			-4.847***
				{0.060}			{0.005}
	Weighted average density – MSOA (log)2			0.336**			0.325***
				{0.039}			{0.002}
Urban rainfall	Urban rainfall per sewer length (log)				0.132***	0.153***	0.152***
					{0.001}	{0.000}	{0.000}
Constant	Constant	-7.956***	-6.609***	-7.572***	-7.809***	-6.424***	-7.492***
		{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Model robustness tests and additional information							
Statistical diagnostic tests	Adjusted R-squared	0.919	0.897	0.895	0.928	0.918	0.916
	RESET test	0.006	0.326	0.399	0.005	0.577	0.824
	VIF (max)	2.337	1.914	1.996	2.53	1.918	2.003
	Pooling / Chow Test	0.9	0.982	0.987	0.965	0.97	0.981
	LM test (Pooled OLS vs RE)	0	0	0	0	0	0
	Normality of model residuals	0.248	0.244	0.376	0.007	0.002	0.005
	Heteroskedasticity of model residuals	0.183	0.034	0.027	0.04	0.002	0.001
Model information	Estimation method	RE	RE	RE	RE	RE	RE
	Observations	110	110	110	110	110	110
	Dependent variable	Sewage collection botex plus					
Efficiency score distribution	Minimum	0.91	0.88	0.86	0.92	0.89	0.87
	Maximum	1.08	1.21	1.16	1.09	1.18	1.13
	Range	0.17	0.33	0.30	0.16	0.29	0.26
Sensitivity tests	Removal most efficient company	A	A	A	A	G	A
	Removal least efficient company	A	G	A	A	G	G
	Removal first year	A	G	G	A	G	A
	Removal last year	G	G	A	A	G	G

The consultation mentions that the version of model SWC1 indicates "*that the squared density term is strongly insignificant when properties per sewer length is used as the density variable*"[25]. We disagree that the squared term is strongly insignificant. In fact, it is significant within the range of 15% significance level (see model SWC1_SQ in table 2). This level of significance is consistent with the

[25] See Econometric Base Cost Models for PR24 Consultation, p. 43.

November 2022 Ofwat's template guidance for the base econometric model submission that states[26]:

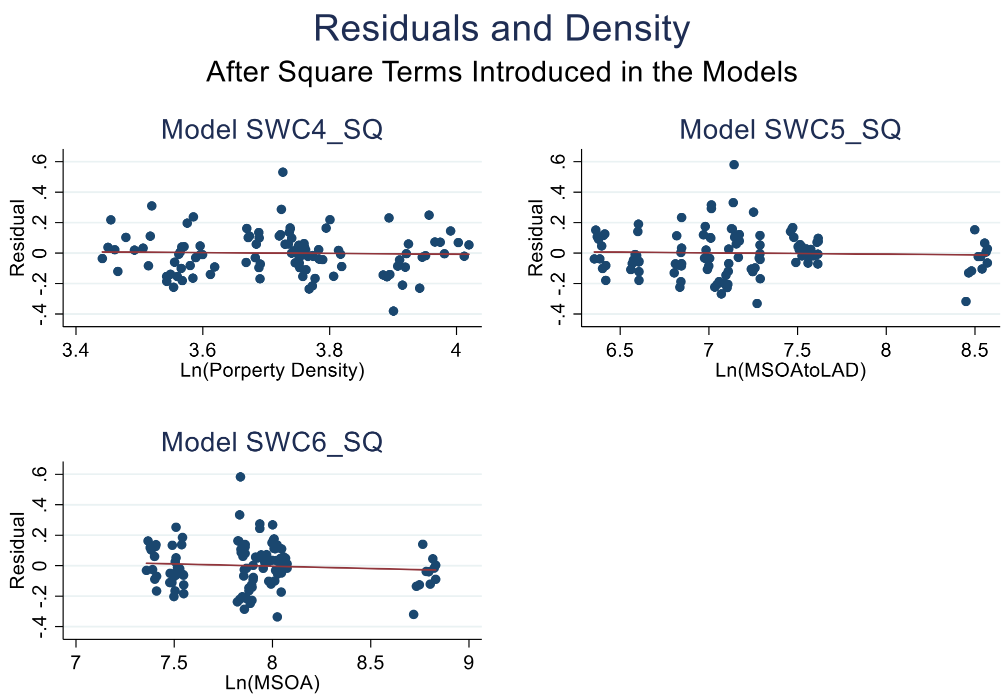
- "*A higher p-value indicates a lower level of statistical significance (ie there is less confidence in the value of the estimated coefficient). However, there is a wide range of confidence levels in this category. Statistical significance of 80% and even 70% may be deemed valid in practical work*" (p. 17).
- Furthermore, the consultation document does not recognise that the square density term becomes statistically significant when the urban rainfall rate is added in model SWC4_SQ in table 2.

The sensitivity tests presented in table 2 for all the models suggest different ambers that we would like to briefly explain. Having an amber really does not undermine the results of the models, as the changes that are presented with these sensitivity tests are marginal. In most of the cases some drivers become more or sometimes less statistically significant and in a few cases some drivers became not statistically significant at 10% or lower, but this could be expected in a very small sample of only 10 cross-sectional observations (N=10). There is not a substantial or systematic change of coefficient signs, or and overall underperformance of the models. With this in mind we think that the robustness of the models with the square term of density is strong.

These econometric results suggest that there is clear empirical evidence to include the square term of density. These econometric results support the economic rationale presented in the previous section. In figure 5, we calculate the residuals of the models that use the nonlinear effect of density and found that the U-shape relationship with density is dissipated. This provides more evidence to support the inclusion of the square term of density.

[26] https://www.ofwat.gov.uk/wp-content/uploads/2022/11/Template_and_guidance_for_model_submission.pdf

Figure 5

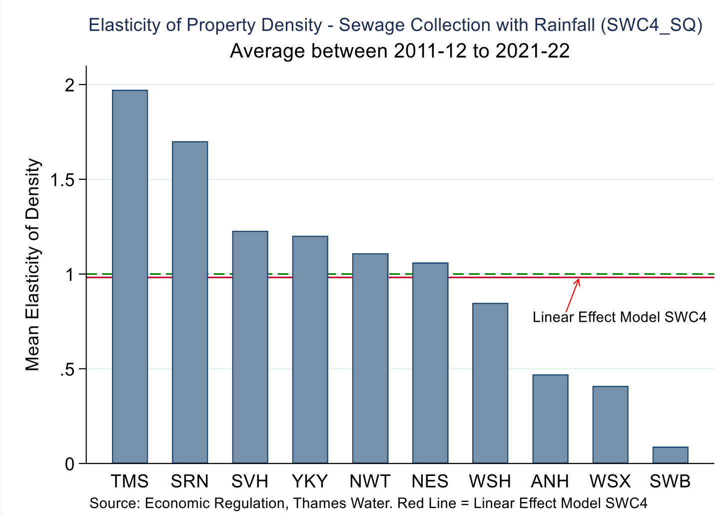


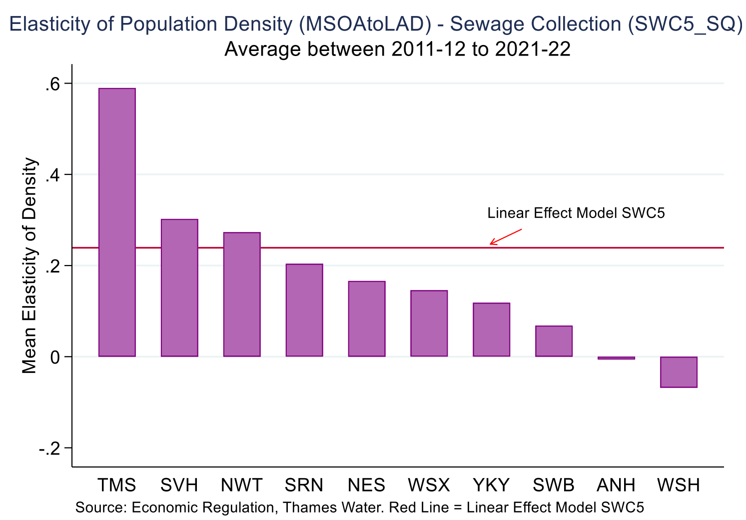
Source: Economic Regulation, Thames Water.

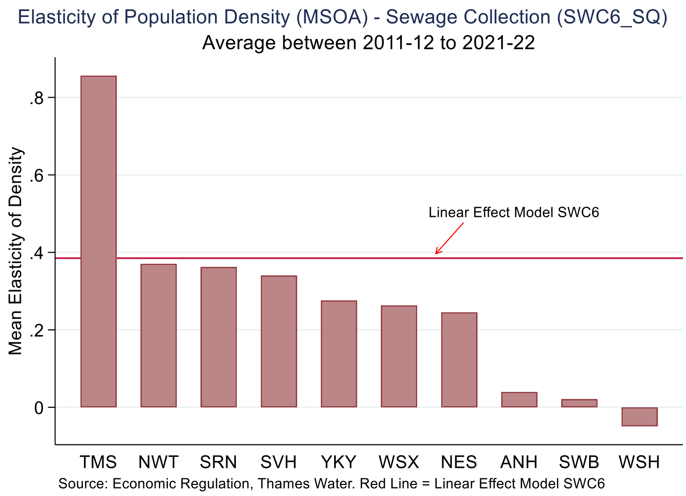
Elasticity of Density in Sewage Collection

Figure 6 illustrates the elasticity of density that is derived from each of the models in sewage collection SWC4_SQ, SWC5_SQ and SWC6_SQ. Similar results are found when rainfall is excluded from the models.

Figure 6







To calculate the elasticity of density for each company per year (ϵ_{D_it}) we use the expression below derived from the SWC models that include rainfall rates and the square term of density:

$$\epsilon_{D\_it} = \frac{\partial Ln(Botex_{it})}{\partial Ln(Density_{it})} = \beta_{Density} + 2\beta_{Density\_SQ} Ln(Density_{it})$$

These elasticities are depicted in figure 6 for all the three versions of density. In the charts of figure 6 we add the linear effects of models SWC4, SWC5 and SWC6 proposed in the consultation with a horizontal red line. We notice that there are few companies that are reflected on this average but most of the industry distant from this effect.

These graphs suggest that there are significant differences regarding the cost faced by companies across the industry and the different levels of density they face, and a linear effect is not representative of what companies are facing in the industry. A model that omits the squared term would significantly distort the efficiency assessment and, consequently, significantly overfund companies with low densities and underfund companies with high densities.

<u>Functional Form Test: Restricted (Linear) versus Unrestricted (Nonlinear)</u>

Finally, we test the functional form of the *restricted* models that do not include the square term of density against the *unrestricted* or more flexible models that include the square term of density.

We follow Greene (2012)[27] approach in testing these restrictions. All models reject the null hypothesis Ho: square term effect null at 10% and all the models at 5% except for SWC1 (see table 3 for the P-values of the F-test). This Functional Form test between restricted and unrestricted models provides more evidence that a more flexible functional form (square term of density) is preferred to the restricted version of the models presented in the consultation.

Table 3

Restricted (Linear) versus Unrestricted (Nonlinear) F-test

	SWC1 v SWC1_SQ	SWC2 v SWC2_SQ	SWC3 v SWC3_SQ	SWC4 v SWC4_SQ	SWC5 v SWC5_SQ	SWC6 v SWC6_SQ
P-value	0.0619	0.0017	0.0060	0.0000	0.0006	0.0010

Sources: Economic Regulation, Thames Water
Note: All resutls are based on OLS outputs.

Each column in table 3 represents the restricted model proposed in the consultation (SWC) against the unrestricted model (SWC_SQ) that includes the square term of density. The economic and econometric evidence above strongly supports the inclusion of a square term of density in all the sewage collection models.

Q4.5) Do you agree with the inclusion of urban rainfall in our sewage collection and wastewater network plus models?

We agree with the inclusion of the urban rainfall in SWC and WWWNP models. The inclusion goes in line with our suggestion in the PR24 base consultation response in December 2021 and with our proposed model's submission in January 2023.

There is strong evidence on the improvement of the SWC and WWWNP models. Excluding the driver could lead us to potential omitted variable biases issues. Moreover, all models' R^2 improve significantly with the inclusion of the driver.

We recognise that including the urban annual rainfall in the models decrease Thames Water (TMS) efficiency position. However, we think that including the urban annual rainfall in the models is appropriate for the mid- and long-term stability of the industry regarding the effect of variables related to climate change pressures.

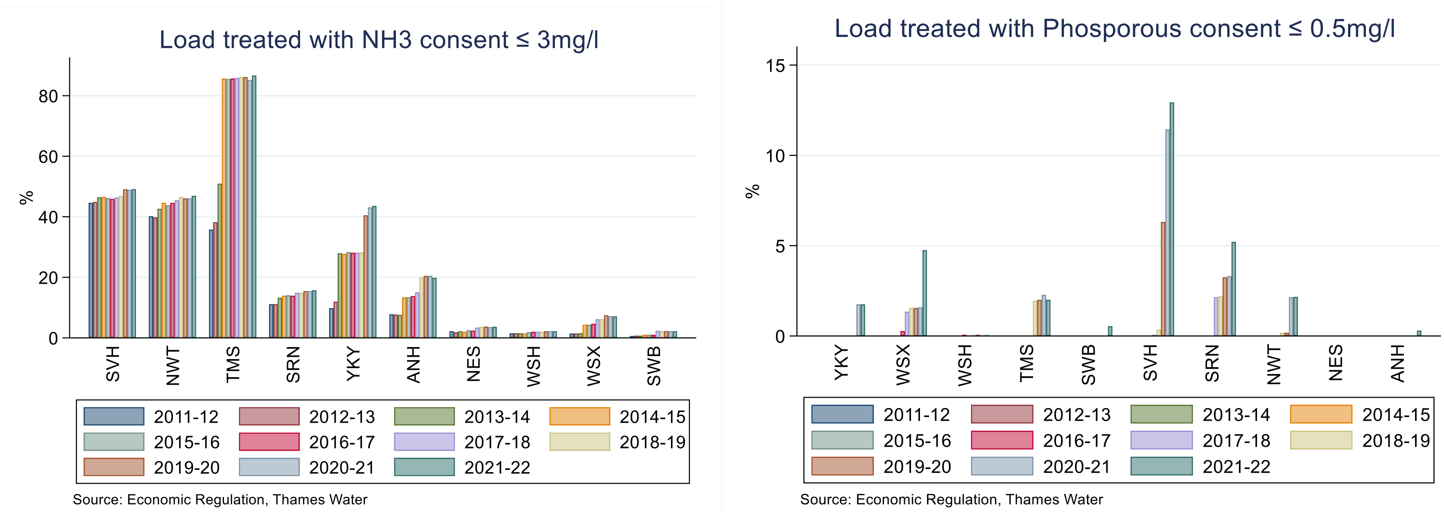
[27] See Greene (2012) Econometric Analysis, Chapter 5, Seventh Edition, International Edition.

Q4.6) Do you agree with our approach to capturing sewage treatment complexity in our proposed wastewater network plus base cost models?

We agree with the use of Load treated with ammonia consent ≤ 3mg/l as the variable that captures the sewage treatment complexity across the industry. The variable shows sufficient variability (see figure 7) and is known to be a material driver of costs.

The driver is robust to sensitive tests and highly significant across all model specifications. The stability of the driver is also significant to highlight. We see that Ammonia consent 3ml/g has a strong representation of treatment complexity in base costs for SWT and WWWNP models.

Figure 7



What are your views on our proposed options to account for additional ongoing cost associated with P-removal?

g. Models with a P-driver (eg percentage of load with a P-permit <= 0.5mg/l) fixed at the 2024/25 level.

h. A post-modelling adjustment that funds efficient ongoing opex associated with P-removal using data provided by companies in APRs.

i. Cost adjustment claims.

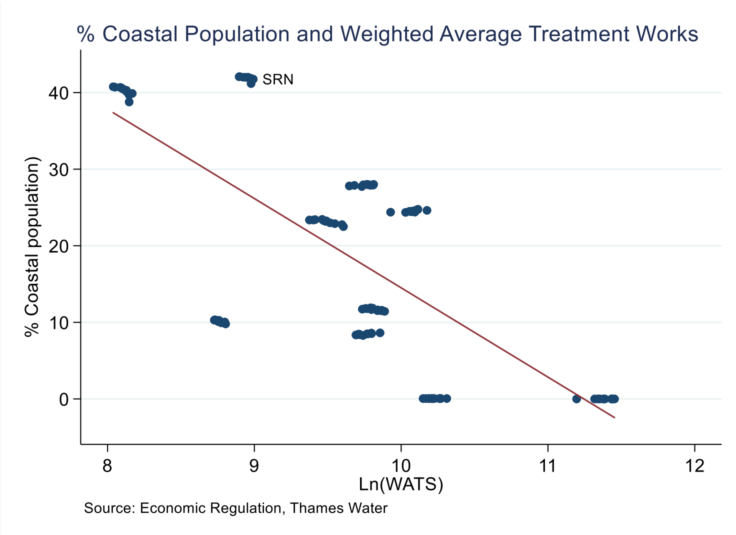
Taking an approach from the modelling point of view would not reflect the cost structure that companies face, and perhaps a CAC case is more suitable as companies face the p-removal restrictions at different rates.

Q4.7) Do you agree with Southern Water's proposal to include the percentage of population living in coastal areas in sewage treatment models?

We disagree. We think that the driver might be captured in a certain degree by the inclusion of WATS or any of the economies of scale drivers used at SWT models. For example, a highly densely populated area closer to a coastal line would be linked to large sewage treatment works.

WATS or any of the economies of scale drivers used in the SWT models captures until some degree the contribution of each area. In the WATS case the disaggregation of SWTs above band 5 would capture areas highly dense or closer to a coastal area. However, by not having a full contribution in lower bands, might affect the correlation for some companies like the case of SRN. This could be an example, of why the WATS need to be more consistent across the methods used to measure it (see figure 8).

Figure 8



Population density (MOSA) is highly correlated with WATS (0.75). In particular, MSOA reflect this link stronger than the other density variables. This could reinforce the use of MSOA density across the wastewater network.

We run a set of basic Random Effects panel models that explain the correlation between each of the economies of scale drivers and the coastal area driver (see table 4).

Table 4

Economies of Scale at SWT and Coastal Areas

	Coast1 b/se	Coast2 b/se	Coast3 b/se
Coastal_Area	-0.050***	-0.972***	0.148**
	(0.015)	(0.328)	(0.063)
constant	10.579***	72.169***	1.077
	(0.355)	(6.433)	(0.899)
R2_Overall	0.476	0.449	0.348
Wald_Chi2	11.515	8.792	5.582
RESET_P_value	0.85	0.42	0.63
BPagan_Test_P_value	0.00	0.00	0.00
N_Sample_Size	10.00	10.00	10.00
T_Sample_Size	11.00	11.00	11.00
Observations	110.00	110.00	110.00

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

The models suggest a fairly R^2. This means that Coastal areas are reflected until some degree with the economies of scale drivers in SWT.

Given that this variable mainly relates to SRN, and concerns with the stability of the variable's coefficient when SRN is removed, we consider that this driver proposed by SRN is more related to a CAC case.

Bioresources Questions

Q5.1) Do you agree with our proposed set of bioresources cost models?

We welcome the inclusion of the MSOA density driver in the models. However, we do not agree with the proposed set of bioresources cost models on account of the omission of transportation costs and the squared term of density from the drivers.

Transportation Costs

For PR24, it is our opinion that the bioresources costs models incorporates all elements of the sludge treatment, transport, and storage (but excludes quality). We feel that the costs associated with the transportation of the end product to its final location has not been adequately captured.

We expressed this opinion to Ofwat through the model's submission process and Ofwat did not include this driver in the model due to the argument that the location of the Sludge Treatment Centres (STC's) is under "Management Control", making the driver inconsistent with the cost assessment principles.

We disagree with this argument as companies' only have a limited choice over where a Sludge Treatment Centre is located. The efficient and practical choice is to locate these STC's on Treatment Works where large volumes of sewage sludge are generated. This is because it is better to keep raw sewage sludge within the confines of the STW and because raw sewage sludge is more expensive to transport than treated sludge. In addition, transporting raw sludge would result in increased costs due to the greater volume of sludge being transported, increased vehicle movements causing nuisance to customers and the wider road users, and increased carbon emissions due to these vehicles among others. There is a strong incentive on companies to minimise costs, so companies will locate STCs in the best possible location, which will, in the vast majority of cases, be close to major STWs, the location of which is largely driven by location of population centres. So, the suggestion that companies have management control over the location of STCs is seriously flawed.

Whether a Sludge Treatment Centre (STC) is located as close to the farmland or reclamation sites as possible, or whether it is located as close as possible to the where the sludge is, there is a transportation cost incurred, which feeds into the total cost, and it is necessary to reflect this in the cost drivers. There is a strong positive Pearson correlation (0.88) between the real Bioresources Botex and the "Total work done in sludge disposal operations carried out by truck" at the levels, which is the proxy proposed for the transport costs.

Below, we show the robustness of the effect of the transport driver on Botex Sludge Transport costs. As seen in table 1 in total cost models TWBRT1 to TWBRT1_least, the coefficient of the driver is positive and significant i.e., Ln(Dry_Sld_Trsp)[28], and it is robust to the removal of different years, as well as the most efficient and least efficient companies from the bioresources models as a sensitivity check.

Table 1

BR Sludge Transport Total Cost Models: Transport Driver Robustness

	TWBRT1 b/se	TWBRT1_2022 b/se	TWBRT1_2021 b/se	TWBRT1_2012 b/se	TWBRT1_2013 b/se	TWBRT1_most b/se	TWBRT1_least b/se
Ln(Sludge_prod)	-0.026	0.030	0.064	-0.067	-0.044	-0.029	-0.005
	(0.183)	(0.173)	(0.168)	(0.193)	(0.223)	(0.191)	(0.183)
Ln(Dry_Sld_Trsp)	0.259***	0.258***	0.267***	0.277***	0.298***	0.260***	0.297***
	(0.076)	(0.058)	(0.063)	(0.090)	(0.102)	(0.081)	(0.086)
constant	0.011	-0.240	-0.476	0.074	-0.202	0.057	-0.395
	(0.895)	(0.878)	(0.950)	(0.931)	(0.899)	(0.990)	(1.002)
R2_Overall	0.392	0.421	0.422	0.369	0.394	0.319	0.410
RESET_P_value	0.417	0.475	0.492	0.395	0.417	0.415	0.456
Observations	110.00	100.00	90.00	100.00	90.00	99.00	99.00

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

The stability of the sign of the driver is also robust when we consider unit cost models TWBRT2 to TWBRT2_least as seen in table 2, where the driver remains positive with the same sensitivity checks as table 1.

Table 2

BR Sludge Transport Unit Cost Models: Transport Driver Robustness

	TWBRT2 b/se	TWBRT2_2022 b/se	TWBRT2_2021 b/se	TWBRT2_2012 b/se	TWBRT2_2013 b/se	TWBRT2_most b/se	TWBRT2_least b/se
Ln(Dry_Sld_Trsp)	0.135	0.145**	0.145**	0.124	0.122	0.141	0.205**
	(0.089)	(0.068)	(0.068)	(0.115)	(0.129)	(0.097)	(0.082)
constant	-3.892***	-3.968***	-3.971***	-3.797***	-3.773***	-3.952***	-4.512***
	(0.766)	(0.577)	(0.528)	(0.987)	(1.099)	(0.855)	(0.711)
R2_Overall	0.118	0.116	0.121	0.121	0.114	0.174	0.098
RESET_P_value	0.382	0.384	0.384	0.381	0.381	0.383	0.396
Observations	110.00	100.00	90.00	100.00	90.00	99.00	99.00

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

We also include the scale variable i.e., Ln(Sludge_prod) in the unit cost models TWBRT3 to TWBRT3_least as seen in table 3 and again, a positive correlation of the driver is

[28] Ln(Dry_Sld_Trsp)=Ln(Total work done in sludge disposal operations carried out by truck)

found. After controlling for scale, the effect is robust from the statistical significance point of view, the sign remains stable through a range of robustness checks as well as its magnitude as presented in table 1 and 2.

Table 3

BR Sludge Transport Unit Cost Models: Transport Driver Robustness

	TWBRT3 b/se	TWBRT3_2022 b/se	TWBRT3_2021 b/se	TWBRT3_2012 b/se	TWBRT3_2013 b/se	TWBRT3_most b/se	TWBRT3_least b/se
Ln(Sludge_prod)	-1.026***	-0.970***	-0.936***	-1.067***	-1.044***	-1.029***	-1.005***
	(0.183)	(0.173)	(0.168)	(0.193)	(0.223)	(0.191)	(0.183)
Ln(Dry_Sld_Trsp)	0.259***	0.258***	0.267***	0.277***	0.298***	0.260***	0.297***
	(0.076)	(0.058)	(0.063)	(0.090)	(0.102)	(0.081)	(0.086)
constant	0.011	-0.240	-0.476	0.074	-0.202	0.057	-0.395
	(0.895)	(0.878)	(0.950)	(0.931)	(0.899)	(0.990)	(1.002)
R2_Overall	0.270	0.275	0.291	0.283	0.281	0.288	0.258
RESET_P_value	0.528	0.555	0.575	0.525	0.551	0.527	0.566
Observations	110.00	100.00	90.00	100.00	90.00	99.00	99.00

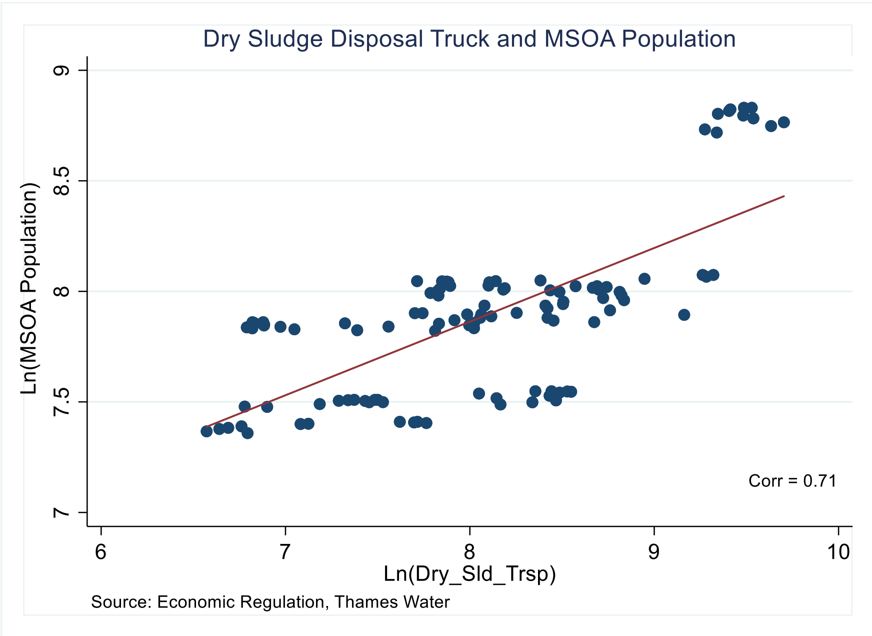
Source: Economic Regulation, Thames Water.
* p<0.10, ** p<0.05, *** p<0.01

Based on the evidence in tables 1 to 3, we have shown the relationship between the transportation driver and sludge transport costs, and we have also shown the stability of the driver regarding its positive effect on costs through several robustness checks. This evidence validates and suggests the inclusion of the driver in the bioresources models.

Density

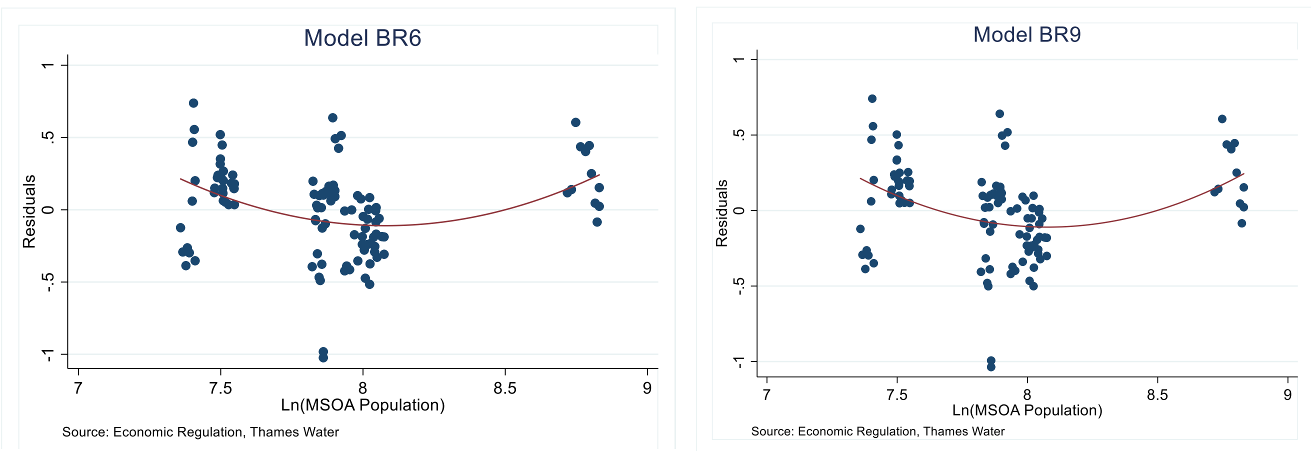
In our submitted models in January 2023, we proposed the inclusion of the squared density term in the bioresources models as companies in areas with higher density will not only receive more sludge but might also face higher costs in terms of transportation of dry sludge to landbanks or landfills. However, Ofwat argues that although the squared term is significant in the models, the relationship might be spurious as there is no noticeable correlation between sludge disposal costs and weighted average density. We disagree with this assessment as companies operating in denser areas have limited availability of farmlands or reclamation sites within close proximity to the STC therefore increasing the distance they have to travel to dispose of dried sludge. This is evidenced by the high correlation between MSOA population density and the “Total measure of 'work' done in sludge disposal operations by truck”, which proxies the transportation costs driver as seen in figure 1.

Figure 1



Also, checking the specification of the models without the squared density term by plotting the residuals of the models against the density variables as seen in figure 2, it suggests that the residuals are dependent on the density variables. Figure 2 shows the residuals from Ofwat models BR6 (Total Cost Model) and BR9 (Unit Cost Model). Both charts in figure 2 show a U-shape relationship between the residuals and density, indicating that at lower and higher levels of density, the residuals are higher than the estimated value. This could be due to the squared term of density being omitted from the models.

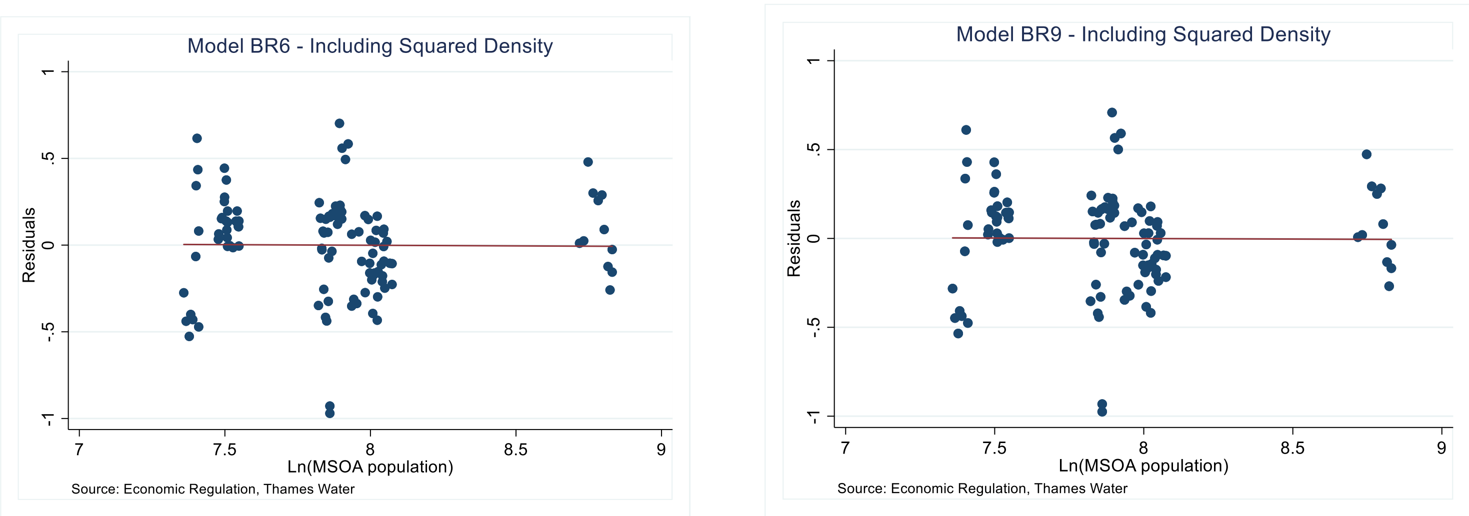
Figure 2



However, when we include the squared density in the same models and plot the residuals against the density driver, the U-shaped relationship is no longer observed. Instead, we observe a linear relationship as seen in figure 3. This strengthens the argument that the squared density term should be included in the models.

Additionally, including the squared term of density improves the fitness of the models as seen in the R^2. Specifically, including the squared density term in models BR8 and BR9 increases the R^2 from 0.12 and 0.10 respectively, to 0.24. On this basis, we strongly recommend that the squared term of density should be considered in the Bioresources models.

Figure 3



Q5.2) Do you agree we should use unit cost models to assess bioresources expenditure?

One of the advantages of using unit cost models is the insight it provides in understanding the economies of scale and scope. The unit cost models proposed by Ofwat omit a scale variable from the drivers, thereby imposing a constant returns to scale assumption. However, imposing constant returns to scale may lead to inaccurate benchmarking. Ofwat argues that this assumption is supported because the coefficient of sludge produced is not statistically significant from zero. We do not agree with this conclusion as there are other factors (such as omitted variables or model misspecification) that can cause the scale driver to be statistically insignificant. Additionally, each model contains only one driver, which we do not think adequately captures the variation in the dependent variable evidenced by the low R^2 from the models (i.e., the highest being 0.24 in model BR7).

The unit cost models need to be improved to adequately reflect the costs faced by companies. In the unit cost models BR7_TMS7 to BR10_TMS10 shown in table 4 below, we include the squared density term in BR8_TMS8 and BR9_TMS9, and we include the driver proxying transportation costs in all models. The coefficient of the scale driver i.e. Ln(Sludge_prod) is statistically significant and negative, clearly indicating the presence of economies of scale. This negates the rationale behind Ofwat's imposition of constant returns to scale on the unit cost models. These models also have better fit, seen in the higher R^2, compared to the unit cost models proposed in the consultation. Also, these models pass the RESET test and are also robust to the removal of the first and last year, and the least and most efficient company. So, if unit costs models are to be used to assess bioresources expenditure in PR24, they have to be improved in order to better reflect the costs that companies face.

Table 4

Ofwat BR Models With Squared Density and Transport Driver (Sample: 2011-12 to 2021-22)

	BR7_TMS7 b/se	BR8_TMS8 b/se	BR9_TMS9 b/se	BR10_TMS10 b/se
Ln(Sludge_prod)	-0.263**	-0.372***	-0.324***	-0.188
	(0.129)	(0.095)	(0.075)	(0.177)
pctbands13	0.059***			
	(0.016)			
Ln(Dry_Sld_Trsp)	0.319***	0.372***	0.368***	0.344***
	(0.084)	(0.086)	(0.086)	(0.086)
Ln(WAD_MSOAtoLAD)		-3.901***		
		(0.856)		
Ln(WAD_MSOAtoLAD_SQ)		0.245***		
		(0.059)		
Ln(WAD_MSOA)			-9.080***	
			(2.166)	
Ln(WAD_MSOA_SQ)			0.536***	
			(0.137)	
lnswtwperpro				0.328**
				(0.160)
constant	-2.344***	13.315***	36.003***	0.004
	(0.385)	(3.533)	(8.925)	(0.747)
R2_Overall	0.381	0.483	0.471	0.353
RESET_P_value	0.787	0.247	0.597	0.683
Observations	110.00	110.00	110.00	110.00

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

Retail Questions

Q6.1) Do you agree with our proposed set of residential retail cost models?

We disagree with the proposed set of residential retail cost models. This is because of the removal of transience and proportion of metered customers from the proposed models. These two drivers are important determinants of the efficient costs that companies incur in their operations. We provide more details on our stance under Questions 6.4 and 6.5.

Q6.2) Do you agree with our approach to modelling deprivation, and/or have any views on the selected variables?

We welcome the development in Ofwat's approach to modelling deprivation. We understand that deprivation is a complex issue that is influenced by a variety of economic, social, and environmental factors, such as income (net of housing costs) and employment, and it is difficult to measure. We agree with Ofwat's inclusion of the average number of county court judgements or partial insights accounts per household as another proxy for customer's propensity to default. We agree that this variable is intuitive and adequately captures some aspects of deprivation across England and Wales.

In addition, Ofwat has opted to use the interpolated income score which we also agree with as it performed better in the models compared to the unadjusted income score. However, we are still concerned that due to information on this driver being published every 5 years, the income score deprivation of 2019 is assumed to be the same for the periods 2019-20 and subsequent years pending a new release of the data. This assumption is highly improbable given the Covid-19 pandemic years of 2020 and 2021, and the current cost of living crisis driven by inflationary pressures. Overall, we agree with the approach to modelling deprivation.

Q6.3) Do you agree with the inclusion of Covid-19 dummy variables in the residential retail cost models?

We strongly agree with Ofwat's inclusion of Covid-19 dummy variables in the retail models. Almost all companies increased their bad debt provisions between 2019-21 due to the uncertainty associated with customer's ability to pay because of the pandemic.

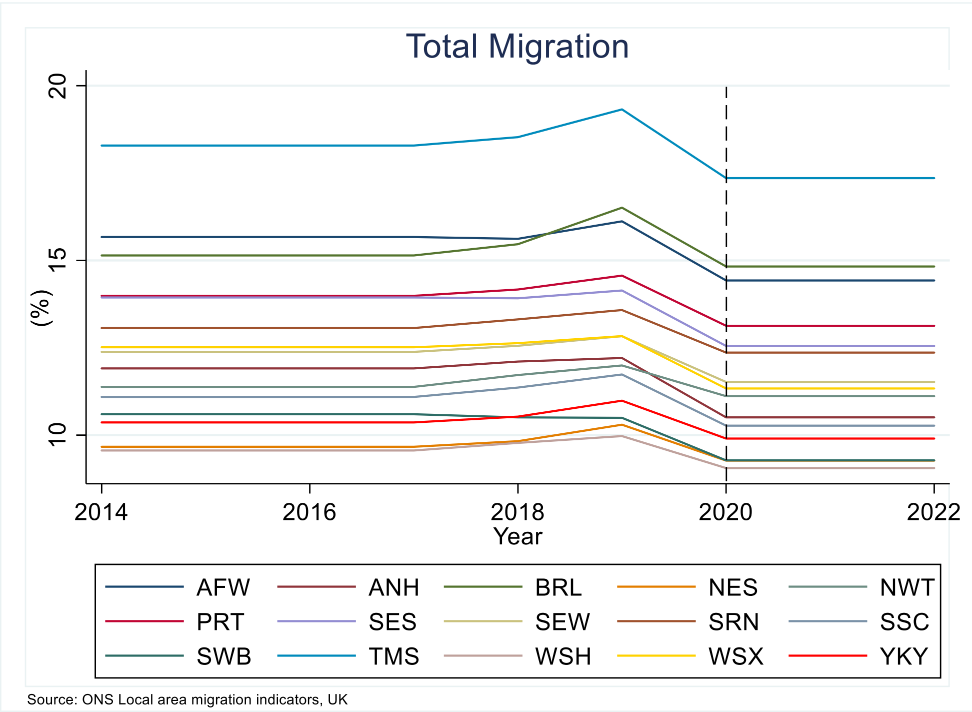
We agree that the use of these dummies improves the performance of the models and mitigates the impact of the pandemic on the cost models.

Q6.4) Do you agree with the removal of transience from the residential retail cost models?

No, we strongly disagree with Ofwat's removal of transience from the residential retail cost models. At PR19, Ofwat believed that transience was a valid driver of retail costs and is outside the control of companies. However, for PR24, Ofwat states that transience does not have a material impact on bad debt costs based on its statistical insignificance, counterintuitive sign, and reduced magnitude of the coefficient in some models. We do not agree with this conclusion as the underlying relationship between transience and costs has not changed.

The reason behind the statistical insignificance, counterintuitive sign, and reduced magnitude of the coefficient of the transience driver in some models is due to the Covid impact. The chart in figure 1 shows a shock to migration for 2019-20 which affected all companies. Migration had previously appeared to be on an upward trend which was reversed by the pandemic in 2020[29].

Figure 1



[29] ONS migration figures are reported mid-year i.e., up to 30 June and so 2019-2020 will include early pandemic effects on Migration.

Once the effect of the pandemic is controlled for in the models through the addition of the Covid 20 and Covid 21 dummies as proposed by Ofwat, the transience driver becomes statistically significant, and the coefficient has the expected sign and magnitude. We demonstrate this in table 1. As seen in models RDC3_TMS3, RTC3_TMS3 and RTC6_TMS6 shown in table 1, the transience driver is positive, statistically significant and has magnitudes comparable to the magnitudes observed in PR19.

Table 1

Ofwat Retail Models With Transience (Sample: 2013-14 to 20121-22)

	RDC1 b/se	RDC2 b/se	RDC3_TMS3 b/se	ROC1 b/se	ROC2 b/se	RTC1 b/se	RTC2 b/se	RTC3_TMS3 b/se	RTC4 b/se	RTC5 b/se	RTC6_TMS6 b/se
Ln(Avg_bill_size)	1.170***	1.207***	1.045***			0.651***	0.659***	0.668***	0.514***	0.540***	0.488***
	(0.119)	(0.132)	(0.107)			(0.085)	(0.089)	(0.096)	(0.066)	(0.068)	(0.070)
Probability_default	0.064***					0.025**			0.021**		
	(0.024)					(0.010)			(0.009)		
Covid20	0.437***	0.395***	0.468***			0.176***	0.153***	0.199***	0.166***	0.147***	0.180***
	(0.101)	(0.090)	(0.105)			(0.030)	(0.029)	(0.033)	(0.030)	(0.030)	(0.033)
Covid21	0.264***	0.193**	0.281***			0.058**	0.026	0.077***	0.044*	0.018	0.055**
	(0.093)	(0.085)	(0.102)			(0.025)	(0.029)	(0.026)	(0.027)	(0.030)	(0.028)
Ln(Avg_Court_judg)		0.879**					0.229			0.181	
		(0.368)					(0.165)			(0.156)	
income_score			0.103***					0.033**			0.031*
			(0.032)					(0.015)			(0.017)
Transience			0.044					0.036***			0.021**
			(0.029)					(0.014)			(0.010)
dual_household				0.002**	0.003***						
				(0.001)	(0.001)						
lnhh_t					-0.045	-0.096***	-0.082***	-0.119***			
					(0.030)	(0.032)	(0.031)	(0.038)			
constant	-5.861***	-5.101***	-5.480***	2.742***	3.324***	0.405	0.626**	0.364	-0.060	0.175	-0.078
	(1.129)	(0.863)	(1.037)	(0.060)	(0.412)	(0.355)	(0.310)	(0.349)	(0.499)	(0.401)	(0.514)
R2_Overall	0.662	0.661	0.692	0.118	0.131	0.697	0.669	0.697	0.650	0.645	0.639
RESET_P_value	0.005	0.012	0.002	0.988	0.312	0.103	0.054	0.212	0.092	0.023	0.206
Observations	153.00	153.00	153.00	153.00	153.00	153.00	153.00	153.00	153.00	153.00	153.00

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

Comparing table 1 and 2, we notice that in models RDC3_TMS3, RTC3_TMS3 and RTC6_TMS6 in table 2 where the Covid dummies are excluded from the bad debt and total cost models, the transience driver then exhibits the counterintuitive signs, statistical insignificance, and reduced magnitudes that Ofwat refers to. So, controlling for the effect of the pandemic through the addition of the Covid dummies also mitigates the issues previously observed with the transience driver. By excluding transience from the models, Ofwat is omitting an important exogenous driver of costs that companies face, and this could bias the estimated results.

Table 2

Ofwat Retail Models With Transience and No Covid Dummies(Sample: 2013-14 to 20121-22)

	RDC1 b/se	RDC2 b/se	RDC3_TMS3 b/se	ROC1 b/se	ROC2 b/se	RTC1 b/se	RTC2 b/se	RTC3_TMS3 b/se	RTC4 b/se	RTC5 b/se	RTC6_TMS6 b/se
Ln(Avg_bill_size)	1.188***	1.188***	1.108***			0.626***	0.629***	0.629***	0.529***	0.536***	0.526***
	(0.128)	(0.132)	(0.109)			(0.087)	(0.089)	(0.095)	(0.062)	(0.068)	(0.065)
Probability_default	0.024					0.008			0.005		
	(0.019)					(0.011)			(0.010)		
Ln(Avg_Court_judg)		0.529					0.108			0.054	
		(0.344)					(0.172)			(0.164)	
income_score			0.044*					0.006			0.006
			(0.026)					(0.016)			(0.017)
Transience			-0.009					0.005			-0.002
			(0.022)					(0.013)			(0.010)
dual_household				0.002**	0.003***						
				(0.001)	(0.001)						
lnhh_t					-0.045	-0.067**	-0.065**	-0.066*			
					(0.030)	(0.033)	(0.032)	(0.036)			
constant	-4.899***	-4.688***	-4.350***	2.742***	3.324***	0.572	0.649**	0.599	0.248	0.302	0.344
	(1.028)	(0.859)	(0.856)	(0.060)	(0.412)	(0.359)	(0.324)	(0.370)	(0.469)	(0.399)	(0.447)
R2_Overall	0.615	0.621	0.626	0.118	0.131	0.634	0.626	0.619	0.614	0.609	0.608
RESET_P_value	0.092	0.083	0.068	0.988	0.312	0.001	0.002	0.004	0.000	0.000	0.000
Observations	153.00	153.00	153.00	153.00	153.00	153.00	153.00	153.00	153.00	153.00	153.00

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

Another reason Ofwat has given for removing the transience driver from the models is the discontinuation of the ONS international migration dataset used to construct the variable. Again, we do not agree that this is a valid reason to remove the driver altogether from the models. Pending an ONS announcement of the replacement for the international migration dataset used to construct this variable and in the absence of any other suitable external data to model transience, for PR24, we think Ofwat can still use this dataset. We propose two ways that this could be done:

Holding the migration rates constant at the 2020 level for the years after 2020: This is also similar to Ofwat's approach to modelling the income deprivation in PR19 where income score was held constant at the 2016-17 rate for the subsequent years. The current Ofwat data for migration uses this approach and as shown in table 1, this yield results that are statistically significant and follows economic intuition.

Extrapolating the migration rates using the growth rate: This is similar to Ofwat's approach to modelling the income deprivation where the interpolated income score is used to measure deprivation as opposed to using the unadjusted income score, which is published every 5 years. Most companies, including TMS, noted that using the unadjusted income score yielded insignificant results and Ofwat agreed with this approach. From models RDC3_TMS3, RTC3_TMS3 and RTC6_TMS6 in table 3, we observe that using the extrapolated migration driver also yields coefficients that are statistically significant, with the right signs and magnitude. Based on these reasons, we reiterate our stance that transience should be included in the models.

Table 3

Ofwat Retail Models With Extrapolated Transience(Sample: 2013-14 to 20121-22)

	RDC1	RDC2	RDC3_TMS3	ROC1	ROC2	RTC1	RTC2	RTC3_TMS3	RTC4	RTC5	RTC6_TMS6
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
Ln(Avg_bill_size)	1.170***	1.207***	1.040***			0.651***	0.659***	0.608***	0.514***	0.540***	0.484***
	(0.119)	(0.132)	(0.100)			(0.085)	(0.089)	(0.084)	(0.066)	(0.068)	(0.071)
Probability_default	0.064***					0.025**			0.021**		
	(0.024)					(0.010)			(0.009)		
Covid20	0.437***	0.395***	0.435***			0.176***	0.153***	0.165***	0.166***	0.147***	0.161***
	(0.101)	(0.090)	(0.099)			(0.030)	(0.029)	(0.030)	(0.030)	(0.030)	(0.031)
Covid21	0.264***	0.193**	0.299**			0.058**	0.026	0.065**	0.044*	0.018	0.058**
	(0.093)	(0.085)	(0.118)			(0.025)	(0.029)	(0.028)	(0.027)	(0.030)	(0.029)
Ln(Avg_Court_judg)		0.879**					0.229			0.181	
		(0.368)					(0.165)			(0.156)	
income_score			0.090***					0.022			0.022
			(0.029)					(0.014)			(0.016)
Transience			0.037					0.018**			0.017**
			(0.027)					(0.008)			(0.007)
dual_household				0.002**	0.003***						
				(0.001)	(0.001)						
lnhh_t					-0.045	-0.096***	-0.082***	-0.081***			
					(0.030)	(0.032)	(0.031)	(0.029)			
constant	-5.861***	-5.101***	-5.193***	2.742***	3.324***	0.405	0.626**	0.552*	-0.060	0.175	0.123
	(1.129)	(0.863)	(0.881)	(0.060)	(0.412)	(0.355)	(0.310)	(0.335)	(0.499)	(0.401)	(0.449)
R2_Overall	0.662	0.661	0.697	0.118	0.131	0.697	0.669	0.685	0.650	0.645	0.649
RESET_P_value	0.005	0.012	0.001	0.988	0.312	0.103	0.054	0.342	0.092	0.023	0.206
Observations	153.00	153.00	153.00	153.00	153.00	153.00	153.00	153.00	153.00	153.00	153.00

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

Q6.5) Do you agree with the removal of 'proportion of metered customers' from the residential retail cost models?

No, we disagree that the proportion of metered customers should be removed from the models. Serving metered customers comes at a higher cost compared to unmeasured customers. First, there are meter reading costs associated with metered customers that do not apply to unmeasured customers. Also, companies with higher meter penetration are likely to receive more customer contacts which can drive increased costs. For instance, based on the number of service tickets raised in 2022-23 period for TMS, billing and charging contacts per metered property were 52% more than that for unmeasured properties. In the same period, billing and charging complaints per metered property were 119% more than that for unmeasured properties. In the report for Ofwat, PwC reiterated the higher costs that come with serving metered customers can be through higher contacts due to the increased variability of bill values or due to estimated bills, and through meter reading costs.[30] This increases the costs to serve metered customers and should be reflected in the models.

Ofwat states that the driver produces highly statistically insignificant results with estimated coefficients close to zero and concludes that meter reading does not have a material impact on retailers' costs. We also observe that this driver does not perform as expected in the other retail costs models. However, table 4 shows that

[30] Retail services efficiency review 2022: Report for Ofwat

this driver performs well in the Total Retail Costs models as seen in RTC1_TMS1 and RTC4_TMS4. Including the proportion of metered customers in two of Ofwat's Total Retail Costs models i.e., RTC1 and RTC4, yields coefficients with magnitudes comparable to PR19, with the correct signs and statistically significant at 16% level. The models are also robust to the removal of the first and last year from the sample, and also the removal of the least and most efficient company. Consequently, we are of the view that the proportion of metered customers should be included in the Total Cost Retail Models.

Table 4

Ofwat Models With Metered Connections

	RTC1_TMS1 b/se	RTC4_TMS4 b/se
Ln(Avg_bill_size)	0.642***	0.495***
	(0.082)	(0.072)
Probability_default	0.045***	0.031**
	(0.015)	(0.014)
metered_connection	0.005	0.002
	(0.003)	(0.003)
lnhh_t	-0.118***	
	(0.030)	
Covid20	0.182***	0.168***
	(0.030)	(0.031)
Covid21	0.062**	0.045*
	(0.026)	(0.027)
constant	0.036	-0.334
	(0.428)	(0.555)
R2_Overall	0.695	0.642
RESET_P_value	0.286	0.150
Observations	153.00	153.00

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