12 January 2023

Cost assessment team
Ofwat

BY EMAIL

Dear Ofwat,

RE: Submission of base cost econometric models ahead of spring 2023 consultation

Thank you for the opportunity to make this submission ahead of your spring consultation. We found the datasets, do files and template guidance very helpful for making this submission. We hope that you find our model ideas helpful to the PR24 process.

As we have set out in a paper we prepared for the future ideas lab[1] we believe that significant collective change is needed from both companies and regulators in order to support a healthier asset base for the long-term. This is likely to require some material change to the way that the efficient cost allowances are set at future price controls and correspondingly the cost models. However, as we have highlighted in the same paper this will be a substantial undertaking that should happen in parallel with other work to establishing and embedding common frameworks for assessing asset management effectiveness in companies and also new and more comprehensive measures of asset health. Given this we still consider that we will need more time to work towards these changes across the sector and perhaps they could be available for PR29.

For PR24 we agree with Ofwat that there does not need to be significant change from the PR19 econometric models. Whilst these models are relatively simple they continue to perform well against statistical performance tests, have good explanatory power and the intuition and engineering logic behind them is sound. This is no doubt why the CMA supported their use in the PR19 appeals and they are also now well understood by the sector having been used previously. We also consider that frequent changes to the models are unhelpful from a regulatory stability point of view as the fundamental engineering and economics of base activities has not changed since PR19.

Given the materiality of the enhancement programme for PR24, we also think time is better spent there in developing cost models for considering a large ramp in expenditure which will be more beneficial for customers and more proportionate overall for PR24 as whole.

We therefore make limited suggestions where we think improvements can be made to better promote the right incentives and where there is improved data available on a cost driver:

[1] See: https://www.ofwat.gov.uk/wp-content/uploads/2022/06/Northumbrian_Water_Resilient_efficient_services_require_healthy_assets.pdf

Northumbrian Water Limited
Registered in England and Wales No 2366703
Registered office: Northumbria House,
Abbey Road, Pity Me, Durham, DH1 5FJ

- **Bioresources**: the current models include variables that are within management control in the medium term and result in different allowances (£/ttds) which does not best promote efficient behaviour nor the facilitation of the bioresources market. We think these models should be simplified. This is discussed in Annex A.
- **Improved weighted average density variables**: We support the use of the updated density variables using Middle Super Output Area (MSOA) population estimates. We agree these mitigate the issue of local authority boundaries changing over time and therefore should be used instead of the Local Authority District data. This is discussed in Annex B.

We have also explored potential models for **Growth at Sewage Treatment Works**. We can replicate the models developed by Arup for Ofwat but do not consider them to be a material improvement over PR19 approach. We think that both approaches are poorly targeted at identifying efficient cost allowances. This is because the costs are driven by local growth and headroom at specific works affected by this. The models are unable to identify these interactions and therefore do not provide robust inferences. We continue to believe that undertaking a deep dive of the DWMP in this area would provide a better and more targeted approach to identifying efficient future growth costs.

As set out above, we don't think a large overhaul of the PR19 models is required and that time would be better spent on the modelling suite for enhancement costs. We look forward to Ofwat's consultation in the spring and would be happy to discuss our response in more detail ahead of this.

Yours sincerely,

Northumbrian Water

Annex A: Bioresoures

The PR19 models in our view have 2 undesirable characteristics in the context of an area where efficient market trading is an opportunity for significant customer benefit:

- They include variables such as 'load treaded in bands 1-3' and 'sewage treatment works per number of properties' which in the medium term are within management control and are part of each companies operational strategy. For example, companies can decide to treat load locally or transport to a larger centralised works. By including these variables, inefficient operating strategies could be rewarded at the expense of customers.
- The models result in allowances which differ significantly between companies on a £/ttds basis. These differences are driven by the variables above and the economies of scale assumed in load and the impact of density. Having these different allowances where companies with high costs could get high allowances does not effectively promote trading with lower cost neighbours and overall market efficiency.

We therefore think the models should be simplified to remove these variables and could even be replaced with a simple cost model that would give each company the same allowance on a £/ttds basis. In developing our thinking we have estimated the models as set out below:

- NESBRENH1 and NESBRENH2 are the BR1 and BR2 models using PR19 cost drivers for comparison purposes.
- NESBRENH3: this removes the load treated in different size band variables which are under management control. When this model is estimated the density variable is no longer statistically significant.
- NESBRENH4: this model further removes the density variable which was not significant (same result if the MSOA version is used) and keeps load as the only cost driver. It is interesting that this model gives a more intuitive coefficient as 0.92 suggests that there are economies of scale whereas the other models suggest diseconomies of scale.

However, assuming economies of scale and allowing different companies different allowances on a £/ttds basis still does not best promote a market as higher cost companies with higher allowances might not be incentivised to trade. We have therefore proposed a simple unit cost model NESBRENH5 which just calculates different companies' unit costs against which an efficiency challenge could be implemented. This would best promote market efficiency and savings for customers as higher cost companies would be incentivised to either cut their cost or trade with more efficient neighbours that can treat sludge on a lower £/ttds basis. It also benefits from reducing complexity in a technical area and regulatory burdens. We therefore suggest that either NESBRENH4 or NESBRENH5 (i.e. a simple unit cost approach) is used.

Model NESBRENH1 (same as PR19 BR1 drivers)

Information requested	Response
Econometric model formula	• NESBRENH1: ln (BR botex plus$_{it}$) = $\alpha + \beta_1$ ln (sludge produced$_{it}$) + β_2 (weighted average density LAD$_{it}$) + β_3 ln (% load treated at levels 1-3$_{it}$) + ε_{it}
Description of the dependent variable:	• Bioresources botex including growth enhancement (code: botex_bio, in Interface_real).
Description of the independent variables:	• Sludge produced (code: sludgeprod), as reported in the published PR24 wholesale dataset. • Weighted average density LAD (code:WAD_LAD), as reported in the published PR24 wholesale dataset • Percentage of load treated at works sized 1-3 (code: pctbands13), as reported in the published PR24 wholesale dataset
Brief comment	The sample covers 11 years of data. There is no difference in the panel structure from the published do file. The model performs relatively well with R2 > 80% The sensitivity tests show that the model is robust when removing the most and least efficient company, and there's a slight change of significance when removing the first and the last year of data.

Model NESBRENH2 (same as PR19 BR2 drivers)

Information requested	Response
Econometric model formula	• NESBRENH2: ln (BR botex$_{it}$) = $\alpha + \beta_1$ ln (sludge produced$_{it}$) + β_2 (number of STWs per property$_{it}$) + ε_{it}
Description of the dependent variable:	• Bioresources botex including growth enhancement (code: botex_bio, in Interface_real)
Description of the independent variables:	• Sludge produced (code: sludgeprod), as reported in the published PR24 wholesale dataset.

	• Number of sewage treatment works per property (code: swtwperpro), as reported in the published PR24 wholesale dataset.
Brief comment	• The sample covers 11 years of data. There is no difference in the panel structure from the published do file. • The model performs relatively well with R2>75% • The sensitivity tests show that the model is still robust even when there are changes to the underlying sample

Model NESBRENH3

Information requested	**Response**
Econometric model formula	• NESBRENH3: $\ln(\text{BR botex}_{it}) = \alpha + \beta_1 \ln(\text{sludge produced}_{it}) + \beta_2 (\text{weighted average density LAD}_{it}) + \varepsilon_{it}$
Description of the dependent variable:	• Bioresources botex including growth enhancement (code: botex_bio, in Interface_real).
Description of the independent variables:	• Sludge produced (code: sludgeprod), as reported in the published PR24 wholesale dataset. • Weighted average density LAD (code: WAD_LAD), as reported in the published PR24 wholesale dataset
Brief comment	• The sample covers 11 years of data. There is no difference in the panel structure from the published do file. • The model performs relatively well with R2>75% • The sensitivity tests show that the model is still robust even when there are changes to the underlying sample

Model NESBRENH4

Information requested	**Response**
Econometric model formula	• **NESBRENH4**: $\ln(\text{BR botex}_{it}) = \alpha + \beta_1 \ln(\text{sludge produced}_{it}) + \varepsilon_{it}$

Description of the dependent variable:	• Bioresources botex including growth enhancement (code: botex_bio, in Interface_real).
Description of the independent variables:	• Sludge produced (code: sludgeprod), as reported in the published PR24 wholesale dataset.
Brief comment	• The sample covers 11 years of data. There is no difference in the panel structure from the published do file. • The model performs relatively well with R2>75% • The sensitivity tests show that the model is still robust even when there are changes to the underlying sample

The model results are shown below in Figure 1 using the guidance and colour coding as set out in the Ofwat template.

Figure 1: Model performance and results for bioresources models

	NESBRENH1	**NESBRENH2**	**NESBRENH3**	**NESBRENH4**
Dependent variable	**Botex bioresources**	**Botex bioresources**	**Botex bioresources**	**Botex bioresources**
Sludge produced (log)	1.184*** {0.000}	1.170*** {0.000}	1.068*** {0.000}	0.921*** {0.000}
Weighted average density LAD (log)	-0.146 {0.282}		-0.245 {0.170}	
Percentage of load treated at levels 1-3 (%)	0.058** {0.043}			
number of STWs per property (log)		0.29 {0.173}		
Constant	-0.84 {0.397}	0.775 {0.358}	0.651 {0.392}	-0.409 {0.481}
Estimation method	RE	RE	RE	RE
N	110	110	110	110
Model robustness test				
R2 adjusted	0.81	0.776	0.77	0.76
RESET test	0.427	0.341	0.072	0.651
VIF(max)	3.086	3.359	2.154	1
Pooling/Chow test	0.103	0.386	0.192	0.284
Normality of model residuals	0.473	0.076	0.077	0.047
Heteroskedasticity of model residuals	0.66	0.943	0.334	0.242

Test of pooled OLS vs RE (LM test)	0	0	0	0
Efficiency score distribution (min & max)	Min:0.6711876 Max:1.422258	Min:0.5980741 Max:1.458161	Min:0.5835308 Max:1.402028	Min:0.5454707 Max:1.593246
Sensitivity of estimated coefficients to removal of most and least efficient company	A	G	G	G
Sensitivity of estimated coefficients to removal of first and last year of the sample	G	G	G	G

The table below shows the results for a simple unit cost approach that could be used to a uniform unit cost allowance across the sector and best promote a bioresources market and efficient decisions by companies. An average unit cost is shown for illustration but could set the could be set at a different level.

Figure 2: NESBRENH5 results using a simple unit cost approach

company	Total cost (last 5 years)	Total sludge (last 5 years)	Unit cost
ANH	431	741	0.58
NES	91	352	0.26
NWT	346	994	0.35
SRN	226	581	0.39
SVH	485	1234	0.39
SWB	119	203	0.58
TMS	689	1829	0.38
WSH	282	374	0.75
WSX	177	338	0.52
YKY	433	733	0.59
Average			0.48

The do file for models NESBRENH1 and NESBRENH2 are called “NESBRENH12”
The do file for models NESBRENH3 and NESBRENH4 are called “NESBRENH34”

Annex B: MSOA density variable

It makes intuitive sense to use the Middle Super Output Area (MSOA) density variable (weighted by population) rather than the local authority district (LAD) variable used at PR19 as it better matches to each company's operating area over time and mitigates the impact of changing local authority boundaries.

We have therefore:

- Re-estimated the PR19 models for water (NESWRP1, NESWRP2, NESTWD1, NESWW1, NESWW2) but have swapped out the LAD measures of density for the MSOA version. These models perform well and give similar results to the LAD version of density as they are similar measures and therefore highly correlated. We think the better measurement of density should lead to more accurate assessments of relative efficiency between companies. We have also presented the PR19 models as a comparison using the LAD variable (NESWRP1_LAD, NESWRP2_LAD, NESTWD1_LAD, NESWW1_LAD, NESWW2_LAD). We find that the coefficients on the density variables are more consistent when using the MSOA density variable across the different levels of aggregation which further supports its use over the LAD version.
- Re-estimated the PR19 model for sewage collection (NESSWC1) but have swapped out the LAD measures of density for the MSOA version. This model also performs well and gives similar results to the LAD version of density as they are similar measures and therefore highly correlated. Again we think the better measurement of density should lead to more accurate assessments of relative efficiency between companies. We have also presented the PR19 models as a comparison using the LAD variable (NESSWC2).

WATER MODELS

Model NESWRP1

Information requested	**Response**
Econometric model formula	• NESWRP1: ln (WRP botex$_{it}$) = α + β_1 ln(properties$_{it}$) + β_2 (% water treated at levels 3-6$_{it}$) + β_3 ln (weighted average density MSOA population$_{it}$) + β_4 (ln (weighted average density MSOA population$_{it}$))^2 + ε_{it}
Description of the dependent variable:	• Water resources plus botex (code: Botex_WRP, in Interface_real), as reported in the published PR24 wholesale dataset

Description of the independent variables:	• Total properties (code: properties), as reported in the published PR24 wholesale dataset. • Percentage of water treated at works sized 3-6 (code: pctwatertreated36), as reported in the published PR24 wholesale dataset. • Weighted average density MSOA weighted by population (code: WAD - MSOA - water - population), as reported in the published PR24 wholesale dataset. • Squared term of weighted average density MSOA population
Brief comment	• The sample covers 11 years of data. There is no difference in the panel structure from the published do file. • The model performs well with R2 >90% • Weighted average density is highly correlated with its squared term which causes VIF score to be high, but this is generally acceptable. • The sensitivity tests show that the model is still robust even when there are changes to the underlying sample.

Model NESWRP2

Information requested	**Response**
Econometric model formula	• NESWRP2: $\ln(\text{WRP botex}_{it}) = \alpha + \beta_1 \ln(\text{properties}_{it}) + \beta_2 \ln(\text{weighted average density MSOA population}_{it}) + \beta_3 (\ln(\text{weighted average density MSOA population}_{it}))^2 + \beta_4 \ln(\text{weighted average complexity}_{it}) + \varepsilon_{it}$
Description of the dependent variable:	• Water resources plus botex (code: Botex_WRP, in Interface_real), as reported in the published PR24 wholesale dataset.
Description of the independent variables:	• Total properties (code: properties), as reported in the published PR24 wholesale dataset. • Weighted average density MSOA weighted by population (code: WAD - MSOA - water - population), as reported in the published PR24 wholesale dataset. • Squared term of weighted average density MSOA population

	• Weighted average level of treatment complexity (code: wac), as reported in the published PR24 wholesale dataset.
Brief comment	• The sample covers 11 years of data. There is no difference in the panel structure from the published do file. • The model performs well with R2 > 85% • Weighted average density is highly correlated with its squared term which causes VIF score to be high, but this is generally acceptable. • The sensitivity tests show the model is robust when removing first and last year of data but there are some changes to significance when the most and least efficient company are removed

Model NESTWD1

Information requested	**Response**
Econometric model formula	• NESTWD1: ln (TWD botex$_{it}$) = α + β_1 ln (weighted average density MSOA population$_{it}$) + β_2 (ln (weighted average density MSOA population$_{it}$))^2 + β_3 ln (length of mains$_{it}$) + β_4 ln (Booster pumping stations per length of mains$_{it}$) + ε_{it}
Description of the dependent variable:	• Treated water distribution botex plus network reinforcement (code: Botex+NR_TWD, in Interface_real), as reported in the published PR24 wholesale dataset.
Description of the independent variables:	• Weighted average density MSOA weighted by population (code: WAD - MSOA - water - population), as reported in the published PR24 wholesale dataset. • Squared term of weighted average density MSOA population • Lengths of main (code: lengthsofmain), as reported in the published PR24 wholesale dataset. • Boosters pumping stations per length of mains (code: boosterperlength), as reported in the published PR24 wholesale dataset

Brief comment	• The sample covers 11 years of data. There is no difference in the panel structure from the published do file. • The model performs well with R2 >90% • Weighted average density is highly correlated with its squared term which causes VIF score to be high, but this is generally acceptable. • The sensitivity tests show that the model is still robust even when there are changes to the underlying sample.

Model NESWW1

Information requested	**Response**
Econometric model formula	• NESWW1: ln (WW botex plus$_{it}$) = α + β_1 ln(properties$_{it}$) + β_2 (% water treated at levels 3-6$_{it}$) + β_3 ln (weighted average density MSOA population$_{it}$) + β_4 (ln (weighted average density MSOA population$_{it}$))^2 + β_5 ln (Booster pumping stations per length of mains$_{it}$) + ε_{it}
Description of the dependent variable:	• Wholesale water botex plus network reinforcement (code: Botex+NR_WW, in Interface_real), as reported in the published PR24 wholesale dataset.
Description of the independent variables:	• Total properties (code: properties), as reported in the published PR24 wholesale dataset. • Percentage of water treated at works sized 3-6 (code: pctwatertreated36), as reported in the published PR24 wholesale dataset. • Weighted average density MSOA weighted by population (code: WAD - MSOA - water - population), as reported in the published PR24 wholesale dataset. • Squared term of weighted average density MSOA population • Boosters pumping stations per length of mains (code: boosterperlength), as reported in the published PR24 wholesale dataset.
Brief comment	• The sample covers 11 years of data. There is no difference in the panel structure from the published do file.

	• Weighted average density is highly correlated with its squared term which causes VIF score to be high, but this is generally acceptable. • The sensitivity tests show the model is robust when removing first and last year of data but there are some changes to significance when the most and least efficient company are removed

Model NESWW2

Information requested	**Response**
Econometric model formula	• NESWW2: ln (WW botex plus$_{it}$) = α + β_1 ln(properties$_{it}$) + β_3 ln (weighted average density MSOA population$_{it}$) + β_4 (ln (weighted average density MSOA population$_{it}$))^2 + β_5 ln (weighted average complexity$_{it}$) + β_6 ln (Booster pumping stations per length of mains$_{it}$) + ε_{it}
Description of the dependent variable:	• Wholesale water botex plus network reinforcement (code: Botex+NR_WW, in Interface_real), as reported in the published PR24 wholesale dataset.
Description of the independent variables:	• Total properties (code: properties), as reported in the published PR24 wholesale dataset. • Weighted average density MSOA weighted by population (code: WAD - MSOA - water - population), as reported in the published PR24 wholesale dataset. • Squared term of weighted average density MSOA population • Booster pumping stations per length of mains (code: boosterperlength), as reported in the published PR24 wholesale dataset. • Weighted average level of treatment complexity (code: wac), as reported in the published PR24 wholesale dataset.
Brief comment	• The sample covers 11 years of data. There is no difference in the panel structure from the published do file.

	• Weighted average density is highly correlated with its squared term which causes VIF score to be high, but this is generally acceptable. • The sensitivity tests show the model is robust when removing first and last year of data but there are some changes to significance when the most and least efficient company are removed.

The model results are shown below in Figure 3 using the guidance and colour coding as set out in the Ofwat template.

Figure 3: Model performance and results for water models using MSOA density

	NESWRP1	NESWRP2	NESTWD1	NESWW1	NESWW2
Dependent variable	Botex water resources plus	Botex water resources plus	Botex treated water distribution	Botex wholesale water	Botex wholesale water
Properties (log)	1.054*** {0.000}	1.057*** {0.000}		1.052*** {0.000}	1.046*** {0.000}
Water treated at levels 3-6 (%)	0.004*** {0.009			0.003** {0.011}	
Weighted average density_MSOA_population (log)	-4.986** {0.017}	-5.048** {0.034}	-5.560*** {0.000}	-4.684*** {0.001}	-4.308*** {0.002}
Weighted average density_MSOA_population squared (log)	0.303** {0.017}	0.306** {0.033}	0.393*** {0.000}	0.301*** {0.000}	0.276*** {0.001}
Weighted average complexity (log)		0.315 {0.234}			0.322** {0.030}
Length of mains (log)			1.026*** {0.000}		
Booster pumping stations per length of mains (log)			0.433*** {0.001}	0.509*** {0.003}	0.486*** {0.003}
Constant	9.416 {0.226}	9.592 {0.286}	15.637*** {0.002}	10.300* {0.056}	8.674 {0.108}
Estimation method	RE	RE	RE	RE	RE
N	187	187	187	187	187
Model robustness test					
R squared	0.901	0.896	0.952	0.963	0.965
RESET test	0.765	0.729	0.122	0.178	0.075
VIF(max)	494.261	514.448	496.846	505.678	526.562
Pooling/Chow test	1	1	0.873	0.987	0.965
Normality of model residuals	0.417	0.416	0.014	0.51	0.574

Heteroskedasticity of model residuals	0	0	0.046	0	0
Test of pooled OLS vs RE (LM test)	0	0	0	0	0
Efficiency score distribution (min & max)	Min: 0.4933967 Max: 1.997069	Min: 0.4727714 Max: 1.983194	Min: 0.7497607 Max: 1.424621	Min: 0.7345613 Max: 1.526707	Min: 0.7366859 Max: 1.531639
Sensitivity of estimated coefficients to removal of most and least efficient company	G	A	G	A	A
Sensitivity of estimated coefficients to removal of first and last year of the sample	G	G	G	G	G

The model results using the PR19 LAD measure of density are presented below for reference only in Figure 4.

Figure 4: Model performance and results for water models using LAD density (as per PR19 models)

	NESWRP3	NESWRP4	NESTWD2	NESWW3	NESWW4
Dependent variable	Botex water resources plus	Botex water resources plus	Botex treated water distribution	Botex wholesale water	Botex wholesale water
Properties (log)	1.074*** {0.000}	1.069*** {0.000}		1.071*** {0.000}	1.059*** {0.000}
Water treated at levels 3-6 (%)	0.006*** {0.000}			0.004*** {0.000}	
Weighted average density_LAD (log)	-1.614*** {0.000}	-1.412*** {0.005}	-2.946*** {0.000}	-2.094*** {0.000}	-1.832*** {0.000}
Weighted average density_LAD (log) squared	0.101*** {0.000}	0.087*** {0.009}	0.235*** {0.000}	0.147*** {0.000}	0.128*** {0.000}
Weighted average complexity (log)		0.377 {0.123}			0.430*** {0.001}
Length of mains (log)			1.077*** {0.000}		
Booster pumping stations per length of mains (log)			0.437*** {0.002}	0.335** {0.032}	0.334** {0.019}
Constant	-5.093*** {0.000}	-5.805*** {0.000}	4.723*** {0.002}	-1.565* {0.074}	-2.589*** {0.001}
Estimation method	RE	RE	RE	RE	RE
N	187	187	187	187	187
Model robustness test					
R squared	0.917	0.907	0.957	0.97	0.971
RESET test	0.439	0.324	0.102	0.223	0.122
VIF(max)	210.27	200.433	211.366	219.181	208.595
Pooling/Chow test	0.995	0.999	0.813	0.869	0.724

Normality of model residuals	0.128	0.574	0.52	0.225	0.838
Heteroskedasticity of model residuals	0	0	0.246	0	0
Test of pooled OLS vs RE (LM test)	0	0	0	0	0
Efficiency score distribution (min & max)	Min: 0.5314376 Max: 2.021829	Min: 0.4998569 Max: 1.979361	Min: 0.7713419 Max: 1.376066	Min: 0.7774395 Max: 1.384858	Min: 0.765841 Max: 1.414317

The do file for water models NESWRP1, NESWRP2, NESTDW1, NESWW1 and NESWW2 are called "NES proposed water models"
The do file for water models NESWRP3, NESWRP4, NESTDW2, NESWW3 and NESWW4 are called "NES WAD_LAD"

WASTEWATER MODELS USING MSOA DENSITY VARIABLE

Model NESSWC1

Information requested	**Response**
Econometric model formula	• NESSWC1: $\ln(\text{SWC botex}_{it}) = \alpha + \beta_1 \ln(\text{total sewer length}_{it}) + \beta_2(\text{pumping capacity per km of sewer}_{it}) + \beta_3 \ln(\text{weighted average density MSOA population}_{it}) + \beta_4(\ln(\text{weighted average density MSOA population}_{it}))^2 + \varepsilon_{it}$
Description of the dependent variable:	• Sewage collection botex (code: botex_sc_sewerflood_rein, in Interface_real), as reported in the published PR24 wholesale dataset
Description of the independent variables:	• Total sewer length (code: sewerlength), as reported in the published PR24 wholesale dataset. • Pumping capacity per km of sewer (code: pumping capacity), as reported in the published PR24 wholesale dataset. • Weighted average density MSOA weighted by population (code: WAD-MSOA-water-population), as reported in the published PR24 wholesale dataset. • Squared term of weighted average density MSOA population
Brief comment	• The sample covers 11 years of data. There is no difference in the panel structure from the published do file.

	• Weighted average density is highly correlated with its squared term which causes VIF score to be high, but this is generally acceptable. • The sensitivity tests show the model is robust when removing first and last year of data but there are some changes to significance when the most and least efficient company are removed.

Model NESSWC2

Information requested	**Response**
Econometric model formula	• NESSWC2: ln (SWC botex$_{it}$) = α + β_1 ln (total sewer length$_{it}$) + β_2 (pumping capacity per km of sewer$_{it}$) + β_3 ln (weighted average density LAD$_{it}$) + β_4 (ln (weighted average density LAD$_{it}$))^2 + ε_{it}
Description of the dependent variable:	• Sewage collection botex (code: botex_sc_sewerflood_rein, in Interface_real), as reported in the published PR24 wholesale dataset
Description of the independent variables:	• Total sewer length (code: sewerlength), as reported in the published PR24 wholesale dataset. • Pumping capacity per km of sewer (code: pumping capacity), as reported in the published PR24 wholesale dataset. • Weighted average density LAD (code: WAD_LAD), as reported in the published PR24 wholesale dataset. • Squared term of weighted average density MSOA population
Brief comment	• The sample covers 11 years of data. There is no difference in the panel structure from the published do file. • Weighted average density is highly correlated with its squared term which causes VIF score to be high, but this is generally acceptable. • The sensitivity tests show the model is robust when removing first and last year of data but there are some changes to significance when the most and least efficient company are removed.

The model results are shown below in Figure 5 using the guidance and colour coding as set out in the Ofwat template.

Figure 5: Model performance and results for sewage collection models

	NESSWC1	**NESSWC2**
Dependent variable	Sewage collection botex plus	Sewage collection botex plus
Sewer length	0.852*** {0.000}	0.859*** {0.000}
pumping capacity per length	0.554*** {0.000}	0.604*** {0.000}
weighted average density MSOA	-5.051* {0.060}	
weighted average density MSOA squared	0.336** {0.039}	
weighted average density LAD		-2.480** {0.021}
weighted average density LAD squared		0.181*** {0.010}
constant	14.241 {0.195}	3.606 {0.395}
Estimation method	RE	RE
N	110	110
Model robustness test		
R2 adjusted	0.895	0.895
RESET test	0.399	0.269
VIF(max)	966.456	399.979
Pooling/Chow test	0.987	0.988
Normality of model residuals	0.376	0.268
Heteroskedasticity of model residuals	0.027	0.051
Test of pooled OLS vs RE (LM test)	0	0
Efficiency score distribution (min & max)	Min:0.8563747 Max:1.157116	Min:0.8740109 Max:1.206244
Sensitivity of estimated coefficients to removal of most and least efficient company	A	A
Sensitivity of estimated coefficients to removal of first and last year of the sample	G	G

The do file for wastewater model NESSWC1 is called “NESSWC1”
The do file for wastewater model NESSWC2 is called “NESSWC2”