Proposed model for Water resources plus density variable

Econometric model formula:

SRNWRP1. $\ln(\text{WW botex water resources plus}_{it}) = \alpha + \beta_1 \ln(\text{properties}_{it}) + \beta_2 (\text{percentage of water treated in complexity levels 3-6}_{it}) + \beta_3 \ln(\text{WAD\_MSOA\_population}_{it}) + \beta_4 \ln(\text{WAD\_MSOA\_population2}_{it}) + \varepsilon_{it}$

SRNWRP2. $\ln(\text{WW botex water resources plus}_{it}) = \alpha + \beta_1 \ln(\text{properties}_{it}) + \beta_2 \ln(\text{water treatment complexity index}_{it}) + \beta_3 \ln(\text{WAD\_MSOA\_population}_{it}) + \beta_4 \ln(\text{WAD\_MSOA\_population2}_{it}) + \varepsilon_{it}$

SRNWRP3. $\ln(\text{WW botex water resources plus}_{it}) = \alpha + \beta_1 \ln(\text{properties}_{it}) + \beta_2 (\text{percentage of water treated in complexity levels 3-6}_{it}) + \beta_3 \ln(\text{WAD\_MSOA\_area}_{it}) + \beta_4 \ln(\text{WAD\_MSOA\_area2}_{it}) + \varepsilon_{it}$

SRNWRP4. $\ln(\text{WW botex water resources plus}_{it}) = \alpha + \beta_1 \ln(\text{properties}_{it}) + \beta_2 \ln(\text{water treatment complexity index}_{it}) + \beta_3 \ln(\text{WAD\_MSOA\_area}_{it}) + \beta_4 \ln(\text{WAD\_MSOA\_area2}_{it}) + \varepsilon_{it}$

SRNWRP5. $\ln(\text{WW botex water resources plus}_{it}) = \alpha + \beta_1 \ln(\text{properties}_{it}) + \beta_2 (\text{percentage of water treated in complexity levels 3-6}_{it}) + \beta_3 \ln(\text{WAD\_MSOAtoLAD\_ population}_{it}) + \beta_4 \ln(\text{WAD\_MSOAtoLAD\_population2}_{it}) + \varepsilon_{it}$

SRNWRP6. $\ln(\text{WW botex water resources plus}_{it}) = \alpha + \beta_1 \ln(\text{properties}_{it}) + \beta_2 \ln(\text{water treatment complexity index}_{it}) + \beta_3 \ln(\text{WAD\_MSOAtoLAD\_ population}_{it}) + \beta_4 \ln(\text{WAD\_MSOAtoLAD\_population2}_{it}) + \varepsilon_{it}$

SRNWRP7. $\ln(\text{WW botex water resources plus}_{it}) = \alpha + \beta_1 \ln(\text{properties}_{it}) + \beta_2 (\text{percentage of water treated in complexity levels 3-6}_{it}) + \beta_3 \ln(\text{WAD\_ MSOAtoLAD\_area}_{it}) + \beta_4 \ln(\text{WAD\_ MSOAtoLAD\_area2}_{it}) + \varepsilon_{it}$

SRNWRP8. $\ln(\text{WW botex water resources plus}_{it}) = \alpha + \beta_1 \ln(\text{properties}_{it}) + \beta_2 \ln(\text{water treatment complexity index}_{it}) + \beta_3 \ln(\text{WAD\_ MSOAtoLAD\_area}_{it}) + \beta_4 \ln(\text{WAD\_ MSOAtoLAD\_area2}_{it}) + \varepsilon_{it}$

Description of the dependent variable

Wholesale water resources plus botex (code: botexwrp), as reported in the published PR24 water base cost Stata do-file.

Description of the explanatory variables

properties: Number of connected properties (code: BN2221 + BN2161), as reported in the published PR24 water base cost Stata do-file.

percentage of water treated in complexity levels 3-6: Percent of water treated in water treatment works with complexity levels 3 to 6 (code: pctwatertreated36), as reported in the published PR24 water base cost Stata do-file.

water treatment complexity index: (code: lnwac), as reported in the published PR24 water base cost Stata do-file.

WAD_MSOA_population: weighted average density calculated from MSOA data, using MSOA population for the weights (code: BN4000), as reported in the published PR24 water base cost Stata do-file.

WAD_MSOA_population2: quadratic (squared) term for the log of the weighted average density calculated from MSOA data, using MSOA population, as reported in the published PR24 water base cost Stata do-file.

WAD_MSOA_area: weighted average density calculated from MSOA data, using MSOA area for the weights (code: BN4001), as reported in the published PR24 water base cost Stata do-file.

WAD_MSOA_area2: quadratic (squared) term for the log of the weighted average density calculated from MSOA data, using MSOA area for the weights, as reported in the published PR24 water base cost Stata do-file.

WAD_MSOAtoLAD_population: weighted average density calculated from MSOA data mapped to LAD level, using LAD population for the weights (code: BN4013), as reported in the published PR24 water base cost Stata do-file.

WAD_MSOAtoLAD_population2: quadratic (squared) term for the log of the weighted average density calculated from MSOA data mapped to LAD level, using LAD population for the weights as reported in the published PR24 water base cost Stata do-file.

WAD_MSOAtoLAD_area: weighted average density calculated from MSOA data mapped to LAD level, using LAD area for the weights (code: BN4014), as reported in the published PR24 water base cost Stata do-file.

WAD_MSOAtoLAD_area2: quadratic (squared) term for the log of the weighted average density calculated from MSOA data mapped to LAD level, using LAD area for the weights, as reported in the published PR24 water base cost Stata do-file.

Brief comment on the models

We propose two sets of four models for consideration. The proposed models are identical to the PR19 models except for the density measure used of which there are two variations.

SRNWRP1 to SRNWRP4

In these four models we replace the Weighted Average Density (WAD) of PR19, which was based on density information at the local authority district (LAD) level, with a WAD that is based on density information at the Middle Layer Super Output Area (MSOA) level.

Ofwat developed the WAD at PR19 to replace the PR14 measure of density. The PR14 density variable measured average (unweighted) density across the entire water company area (or length of pipe). The PR14 measure was neutral to whether an area had a uniform mild density across or whether it had pockets of high and low densities, as long as the total number of customers per square mile was the same. In practice, infrastructure costs between these two types of areas are likely to be different (e.g. due to opportunities to build at scale where there are pockets of density and thereby benefit from economies of scale).

The PR19 WAD was a step in the right direction. It was an improvement on the PR14 density variable.

We consider that MSOA level information may allow a further improvement in the WAD. The MSOA is a smaller geographical unit than the LAD, without being ‘too small’ to capture economies of scale effect on treatment works (MSOAs are between 5,000 and 15,000 people). The MSOA better captures pockets of density and sparsity within a region hence may provide a more accurate information on regional density in water companies’ service areas.

SRNWRP5 to SRNWRP8

In these four models we replace the WAD of PR19 with a WAD that is conceptually the same, but the information about the LAD characteristics (e.g. population, density, and in particular the proportion of the LAD that is attributed to a specific water company) is built up from MSOA level information. That is, there is an MSOA to LAD mapping, and subsequently a LAD to water company mapping.

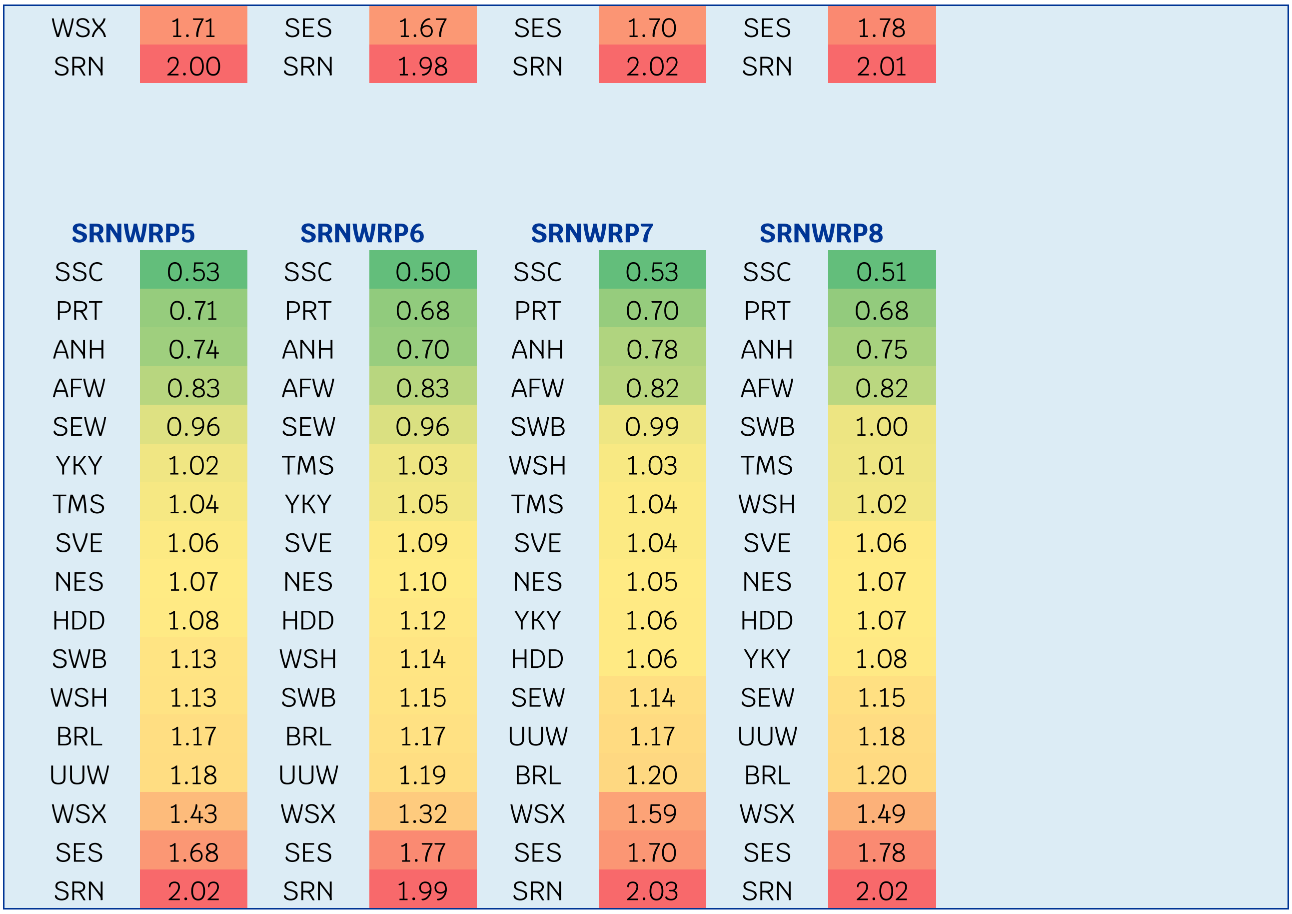
The main advantage of the MSOA to LAD mapping is that the attribution of LADs to water companies is more accurate where a water company straddles across several LADs. This makes the variable more accurate.

	SRNWRP1	**SRNWRP2**	**SRNWRP3**	**SRNWRP4**	**SRNWRP5**	**SRNWRP6**	**SRNWRP7**	**SRNWRP8**
Dependent variable	ln(botex plus water resources $_{it}$)	ln(botex plus water resources $_{it}$)	ln(botex plus water resources $_{it}$)	ln(botex plus water resources $_{it}$)	ln(botex plus water resources $_{it}$)	ln(botex plus water resources $_{it}$)	ln(botex plus water resources $_{it}$)	ln(botex plus water resources $_{it}$)
Properties (log)	1.054*** {0.000}	1.057*** {0.000}	1.082*** {0.000}	1.084*** {0.000}	1.077*** {0.000}	1.075*** {0.000}	1.088*** {0.000}	1.091*** {0.000}
Percentage of water treated in complexity levels 3-6 (%)	0.004*** {0.009}		0.004** {0.014}		0.005*** {0.002}		0.004** {0.012}	
Water treatment complexity index (log)		0.315 {0.234}		0.309 {0.245}		0.343 {0.183}		0.310 {0.242}
WAD_MSOA_population (log)	-4.986** {0.017}	-5.048** {0.034}						
WAD_MSOA_population2 (log)	0.303** {0.017}	0.306** {0.033}						
WAD_MSOA_area (log)			-1.783* {0.063}	-1.854* {0.059}				
WAD_MSOA_area2 (log)			0.135* {0.096}	0.140* {0.085}				
WAD_MSOAtoLAD_population (log)					-1.55*** {0.007}	-1.468** {0.026}		
WAD_MSOAtoLAD_population2 (log)					0.097*** {0.008}	0.091** {0.031}		
WAD_MSOAtoLAD_area (log)							-2.038** {0.030}	-2.111** {0.036}
WAD_MSOAtoLAD_area2 (log)							0.157** {0.048}	0.163* {0.053}
Constant	9.416 {0.226}	9.591 {0.286}	-5.614** {0.011}	-5.527** {0.022}	-5.34*** {0.000}	-5.66*** {0.002}	-4.997** {0.022}	-4.902* {0.051}
Estimation method (OLS or RE)	RE	RE	RE	RE	RE	RE	RE	RE
N (sample size)	187	187	187	187	187	187	187	187
Model robustness tests								

R2 overall	0.90	0.90	0.91	0.90	0.91	0.90	0.91	0.90
RESET test	0.77	0.73	0.47	0.45	0.44	0.67	0.53	0.54
VIF (max)	494	514	307	317	204	204	321	334
Pooling / Chow test	1	1	0.999	0.999	0.999	0.999	0.999	0.999
Normality of model residuals	0.417	0.416	0.362	0.622	0.522	0.812	0.308	0.602
Heteroskedasticity of model residuals	0	0	0	0	0	0	0	0
Test of pooled OLS versus Random Effects (LM test)	0	0	0	0	0	0	0	0
Efficiency score distribution (min and max)	Max: 2.00 Min: 0.49 Gap: 1.50	Max: 1.98 Min: 0.47 Gap: 1.51	Max: 2.02 Min: 0.52 Gap: 1.50	Max: 2.01 Min: 0.50 Gap: 1.51	Max: 2.02 Min: 0.53 Gap: 1.49	Max: 1.99 Min: 0.50 Gap: 1.48	Max: 2.03 Min: 0.53 Gap: 1.51	Max: 2.02 Min: 0.51 Gap: 1.51
Sensitivity of estimated coefficients to removal of most and least efficient company	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**
Sensitivity of estimated coefficients to removal of first and last year of the sample	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**

Efficiency scores distribution

SRNWRP1		**SRNWRP2**		**SRNWRP3**		**SRNWRP4**	
SSC	0.49	SSC	0.47	SSC	0.52	SSC	0.50
PRT	0.66	PRT	0.64	PRT	0.69	PRT	0.67
ANH	0.79	ANH	0.76	ANH	0.78	ANH	0.75
AFW	0.81	AFW	0.82	AFW	0.82	AFW	0.83
TMS	0.92	TMS	0.90	SWB	0.98	SWB	1.00
SEW	1.04	SEW	1.04	WSH	1.03	WSH	1.02
SVE	1.06	SVE	1.09	NES	1.04	TMS	1.04
NES	1.08	NES	1.10	SVE	1.04	NES	1.06
YKY	1.08	SWB	1.11	YKY	1.05	SVE	1.06
SWB	1.09	YKY	1.11	TMS	1.07	YKY	1.07
HDD	1.10	HDD	1.12	HDD	1.07	HDD	1.07
WSH	1.13	WSH	1.13	SEW	1.14	SEW	1.15
BRL	1.15	BRL	1.15	UUW	1.17	UUW	1.18
UUW	1.19	UUW	1.20	BRL	1.18	BRL	1.18
SES	1.59	WSX	1.59	WSX	1.63	WSX	1.53



WSX	1.71	SES	1.67	SES	1.70	SES	1.78
SRN	2.00	SRN	1.98	SRN	2.02	SRN	2.01

SRNWRP5		**SRNWRP6**		**SRNWRP7**		**SRNWRP8**	
SSC	0.53	SSC	0.50	SSC	0.53	SSC	0.51
PRT	0.71	PRT	0.68	PRT	0.70	PRT	0.68
ANH	0.74	ANH	0.70	ANH	0.78	ANH	0.75
AFW	0.83	AFW	0.83	AFW	0.82	AFW	0.82
SEW	0.96	SEW	0.96	SWB	0.99	SWB	1.00
YKY	1.02	TMS	1.03	WSH	1.03	TMS	1.01
TMS	1.04	YKY	1.05	TMS	1.04	WSH	1.02
SVE	1.06	SVE	1.09	SVE	1.04	SVE	1.06
NES	1.07	NES	1.10	NES	1.05	NES	1.07
HDD	1.08	HDD	1.12	YKY	1.06	HDD	1.07
SWB	1.13	WSH	1.14	HDD	1.06	YKY	1.08
WSH	1.13	SWB	1.15	SEW	1.14	SEW	1.15
BRL	1.17	BRL	1.17	UUW	1.17	UUW	1.18
UUW	1.18	UUW	1.19	BRL	1.20	BRL	1.20
WSX	1.43	WSX	1.32	WSX	1.59	WSX	1.49
SES	1.68	SES	1.77	SES	1.70	SES	1.78
SRN	2.02	SRN	1.99	SRN	2.03	SRN	2.02

Proposed model for Treated water distribution density variable

Econometric model formula:

SRNTWD1. ln(WW botex plus treated water distribution$_{it}$) = $\alpha + \beta_1$ ln(lengths of main$_{it}$) + β_2 ln(booster per length$_{it}$) + β_3 ln(WAD_MSOA_population$_{it}$) + β_4 ln(WAD_MSOA_population2$_{it}$) + ε_{it}

SRNTWD2. ln(WW botex plus treated water distribution$_{it}$) = $\alpha + \beta_1$ ln(lengths of main$_{it}$) + β_2 ln(booster per length$_{it}$) + β_3 ln(WAD_MSOA_area$_{it}$) + β_4 ln(WAD_MSOA_area2$_{it}$) + ε_{it}

SRNTWD3. ln(WW botex plus treated water distribution$_{it}$) = $\alpha + \beta_1$ ln(lengths of main$_{it}$) + β_2 ln(booster per length$_{it}$) + β_3 ln(WAD_MSOAtoLAD_ population$_{it}$) + β_4 ln(WAD_MSOAtoLAD_population2$_{it}$) + ε_{it}

SRNTWD4. ln(WW botex plus treated water distribution$_{it}$) = $\alpha + \beta_1$ ln(lengths of main$_{it}$) + β_2 ln(booster per length$_{it}$) + β_3 ln(WAD_ MSOAtoLAD_area$_{it}$) + β_4 ln(WAD_ MSOAtoLAD_area2$_{it}$) + ε_{it}

Description of the dependent variable

Wholesale water treated water distribution botex (code: botexplustwd). That is, botextwd + network reinforcement (B0201DSITDWNC + B0201DSITDWNO), as reported in the published PR24 water base cost Stata do-file.

Description of the explanatory variables

lengths of main: Lengths of main for TWD (code: BN1100), as reported in the published PR24 water base cost Stata do-file.

booster per length: number of booster pumping stations per lengths of main (code: boosterperlength), as reported in the published PR24 water base cost Stata do-file.

WAD_MSOA_population: weighted average density calculated from MSOA data, using MSOA population for the weights (code: BN4000), as reported in the published PR24 water base cost Stata do-file.

WAD_MSOA_population2: quadratic (squared) term for the log of the weighted average density calculated from MSOA data, using MSOA population, as reported in the published PR24 water base cost Stata do-file.

WAD_MSOA_area: weighted average density calculated from MSOA data, using MSOA area for the weights (code: BN4001), as reported in the published PR24 water base cost Stata do-file.

WAD_MSOA_area2: quadratic (squared) term for the log of the weighted average density calculated from MSOA data, using MSOA area for the weights, as reported in the published PR24 water base cost Stata do-file.

WAD_MSOAtoLAD_population: weighted average density calculated from MSOA data mapped to LAD level, using LAD population for the weights (code: BN4013), as reported in the published PR24 water base cost Stata do-file.

WAD_MSOAtoLAD_population2: quadratic (squared) term for the log of the weighted average density calculated from MSOA data mapped to LAD level, using LAD population for the weights as reported in the published PR24 water base cost Stata do-file.

WAD_MSOAtoLAD_area: weighted average density calculated from MSOA data mapped to LAD level, using LAD area for the weights (code: BN4014), as reported in the published PR24 water base cost Stata do-file.

WAD_MSOAtoLAD_area2: quadratic (squared) term for the log of the weighted average density calculated from MSOA data mapped to LAD level, using LAD area for the weights, as reported in the published PR24 water base cost Stata do-file.

Brief comment on the models

We propose two sets of four models for consideration. The proposed models are identical to the PR19 models except for the density measure used of which there are two variations.

SRNTWD1 and SRNTWD2

In these two models we replace the Weighted Average Density (WAD) of PR19, which was based on density information at the local authority district (LAD) level, with a WAD that is based on density information at the Middle Layer Super Output Area (MSOA) level.

Ofwat developed the WAD at PR19 to replace the PR14 measure of density. The PR14 density variable measured average (unweighted) density across the entire water company area (or length of pipe). The PR14 measure was neutral to whether an area had a uniform mild density across or whether it had pockets of high and low densities, as long as the total number of customers per square mile was the same. In practice, infrastructure costs between these two types of areas are likely to be different (e.g. due to opportunities to build at scale where there are pockets of density and thereby benefit from economies of scale).

The PR19 WAD was a step in the right direction. It was an improvement on the PR14 density variable.

We consider that MSOA level information may allow a further improvement in the WAD. The MSOA is a smaller geographical unit than the LAD, without being 'too small' to capture economies of scale effect on treatment works (MSOAs are between 5,000 and 15,000 people). The MSOA better captures pockets of density and sparsity within a region hence may provide a more accurate information on regional density in water companies' service areas.

SRNTWD3 and SRNTWD4

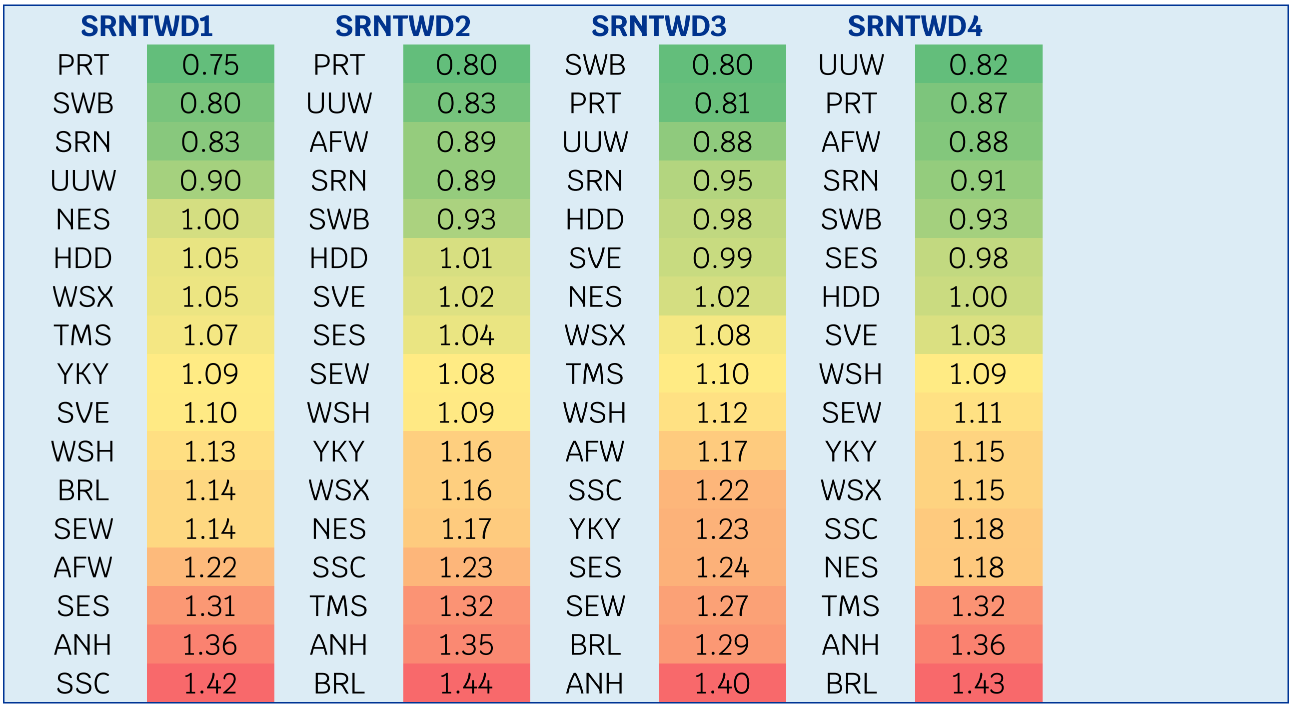
In these two models we replace the WAD of PR19 with a WAD that is conceptually the same, but the information about the LAD characteristics (e.g. population, density, and in particular the proportion of the LAD that is attributed to a specific water company) is built up from MSOA level information. That is, there is an MSOA to LAD mapping, and subsequently a LAD to water company mapping.

The main advantage of the MSOA to LAD mapping is that the attribution of LADs to water companies is more accurate where a water company straddles across several LADs. This makes the variable more accurate.

	SRNTWD1	SRNTWD2	SRNTWD3	SRNTWD4
Dependent variable	ln(botex plus treated water distribution_{it})	ln(botex plus treated water distribution_{it})	ln(botex plus treated water distribution_{it})	ln(botex plus treated water distribution_{it})
Lengths of main (log)	1.026*** {0.000}	1.163*** {0.000}	1.070*** {0.000}	1.163*** {0.000}
Booster per length (log)	0.433*** {0.001}	0.511*** {0.000}	0.461*** {0.002}	0.530*** {0.000}
WAD_MSOA_population (log)	-5.561*** {0.000}			

WAD_MSOA_population2 (log)	0.393*** {0.000}			
WAD_MSOA_area (log)		-3.981*** {0.002}		
WAD_MSOA_area2 (log)		0.380*** {0.001}		
WAD_MSOAtoLAD_population (log)			-2.729*** {0.000}	
WAD_MSOAtoLAD_population2 (log)			0.219*** {0.000}	
WAD_MSOAtoLAD_area (log)				-4.122*** {0.002}
WAD_MSOAtoLAD_area2 (log)				0.395*** {0.001}
Constant	15.638*** {0.002}	5.346* {0.082}	4.155*** {0.008}	5.759* {0.063}
Estimation method (OLS or RE)	RE	RE	RE	RE
N (sample size)	187	187	187	187
Model robustness tests				
R2 overall	0.95	0.96	0.96	0.96
RESET test	0.12	0.01	0.09	0.01
VIF (max)	497	316	207	327
Pooling / Chow test	0.87	0.79	0.80	0.74
Normality of model residuals	0.014	0.053	0.072	0.031
Heteroskedasticity of model residuals	0.046	0.594	0.132	0.990
Test of pooled OLS versus Random Effects (LM test)	0	0	0	0
Efficiency score distribution (min and max)	Max: 1.42 Min: 0.75 Gap: 0.67	Max: 1.44 Min: 0.80 Gap: 0.64	Max: 1.40 Min: 0.80 Gap: 0.60	Max: 1.43 Min: 0.82 Gap: 0.60
Sensitivity of estimated coefficients to removal of most and least efficient company	**G**	**G**	**G**	**G**
Sensitivity of estimated coefficients to removal of first and last year of the sample	**G**	**G**	**G**	**G**

Efficiency scores distribution



SRNTWD1		**SRNTWD2**		**SRNTWD3**		**SRNTWD4**	
PRT	0.75	PRT	0.80	SWB	0.80	UUW	0.82
SWB	0.80	UUW	0.83	PRT	0.81	PRT	0.87
SRN	0.83	AFW	0.89	UUW	0.88	AFW	0.88
UUW	0.90	SRN	0.89	SRN	0.95	SRN	0.91
NES	1.00	SWB	0.93	HDD	0.98	SWB	0.93
HDD	1.05	HDD	1.01	SVE	0.99	SES	0.98
WSX	1.05	SVE	1.02	NES	1.02	HDD	1.00
TMS	1.07	SES	1.04	WSX	1.08	SVE	1.03
YKY	1.09	SEW	1.08	TMS	1.10	WSH	1.09
SVE	1.10	WSH	1.09	WSH	1.12	SEW	1.11
WSH	1.13	YKY	1.16	AFW	1.17	YKY	1.15
BRL	1.14	WSX	1.16	SSC	1.22	WSX	1.15
SEW	1.14	NES	1.17	YKY	1.23	SSC	1.18
AFW	1.22	SSC	1.23	SES	1.24	NES	1.18
SES	1.31	TMS	1.32	SEW	1.27	TMS	1.32
ANH	1.36	ANH	1.35	BRL	1.29	ANH	1.36
SSC	1.42	BRL	1.44	ANH	1.40	BRL	1.43

Proposed model for Wholesale water plus density variable

Econometric model formula:

SRNWW1. ln(WW botex wholesale water plus$_{it}$) = $\alpha + \beta_1$ ln(properties$_{it}$) + β_2 (percentage of water treated in complexity levels 3-6$_{it}$) + β_3 ln(booster per length$_{it}$) + β_4 ln(WAD_MSOA_population$_{it}$) + β_5 ln(WAD_MSOA_population2$_{it}$) + ε_{it}

SRNWW2. ln(WW botex wholesale water plus$_{it}$) = $\alpha + \beta_1$ ln(properties$_{it}$) + β_2 ln(water treatment complexity index$_{it}$) + β_3 ln(booster per length$_{it}$) + β_4 ln(WAD_MSOA_population$_{it}$) + β_5 ln(WAD_MSOA_population2$_{it}$) + ε_{it}

SRNWW3. ln(WW botex wholesale water plus$_{it}$) = $\alpha + \beta_1$ ln(properties$_{it}$) + β_2 (percentage of water treated in complexity levels 3-6$_{it}$) + β_3 ln(booster per length$_{it}$) + β_4 ln(WAD_MSOA_area$_{it}$) + β_5 ln(WAD_MSOA_area2$_{it}$) + ε_{it}

SRNWW4. ln(WW botex wholesale water plus$_{it}$) = $\alpha + \beta_1$ ln(properties$_{it}$) + β_2 ln(water treatment complexity index$_{it}$) + β_3 ln(booster per length$_{it}$) + β_4 ln(WAD_MSOA_area$_{it}$) + β_5 ln(WAD_MSOA_area2$_{it}$) + ε_{it}

SRNWW5. ln(WW botex wholesale water plus$_{it}$) = $\alpha + \beta_1$ ln(properties$_{it}$) + β_2 (percentage of water treated in complexity levels 3-6$_{it}$) + β_3 ln(booster per length$_{it}$) + β_4 ln(WAD_MSOAtoLAD_ population$_{it}$) + β_5 ln(WAD_MSOAtoLAD_population2$_{it}$) + ε_{it}

SRNWW6. ln(WW botex wholesale water plus$_{it}$) = $\alpha + \beta_1$ ln(properties$_{it}$) + β_2 ln(water treatment complexity index$_{it}$) + β_3 ln(booster per length$_{it}$) + β_4 ln(WAD_MSOAtoLAD_population$_{it}$) + β_5 ln(WAD_MSOAtoLAD_population2$_{it}$) + ε_{it}

SRNWW7. ln(WW botex wholesale water plus$_{it}$) = $\alpha + \beta_1$ ln(properties$_{it}$) + β_2 (percentage of water treated in complexity levels 3-6$_{it}$) + β_3 ln(booster per length$_{it}$) + β_4 ln(WAD_MSOAtoLAD_area$_{it}$) + β_5 ln(WAD_ MSOAtoLAD_area2$_{it}$) + ε_{it}

SRNWW8. ln(WW botex wholesale water plus$_{it}$) = $\alpha + \beta_1$ ln(properties$_{it}$) + β_2 ln(water treatment complexity index$_{it}$) + β_3 ln(booster per length$_{it}$) + β_4 ln(WAD_ MSOAtoLAD_area$_{it}$) + β_5 ln(WAD_ MSOAtoLAD_area2$_{it}$) + ε_{it}

Description of the dependent variable

Wholesale water plus botex (code: botexplusww). That is, botexww + network reinforcement (B0201DSITDWNC + B0201DSITDWNO), as reported in the published PR24 water base cost Stata do-file.

Description of the explanatory variables

properties: Number of connected properties (code: BN2221 + BN2161), as reported in the published PR24 water base cost Stata do-file.

percentage of water treated in complexity levels 3-6: Percent of water treated in water treatment works with complexity levels 3 to 6 (code: pctwatertreated36), as reported in the published PR24 water base cost Stata do-file.

water treatment complexity index: (code: lnwac), as reported in the published PR24 water base cost Stata do-file.

booster per length: number of booster pumping stations per lengths of main (code: boosterperlength), as reported in the published PR24 water base cost Stata do-file.

WAD_MSOA_population: weighted average density calculated from MSOA data, using MSOA population for the weights (code: BN4000), as reported in the published PR24 water base cost Stata do-file.

WAD_MSOA_population2: quadratic (squared) term for the log of the weighted average density calculated from MSOA data, using MSOA population, as reported in the published PR24 water base cost Stata do-file.

WAD_MSOA_area: weighted average density calculated from MSOA data, using MSOA area for the weights (code: BN4001), as reported in the published PR24 water base cost Stata do-file.

WAD_MSOA_area2: quadratic (squared) term for the log of the weighted average density calculated from MSOA data, using MSOA area for the weights, as reported in the published PR24 water base cost Stata do-file.

WAD_MSOAtoLAD_population: weighted average density calculated from MSOA data mapped to LAD level, using LAD population for the weights (code: BN4013), as reported in the published PR24 water base cost Stata do-file.

WAD_MSOAtoLAD_population2: quadratic (squared) term for the log of the weighted average density calculated from MSOA data mapped to LAD level, using LAD population for the weights as reported in the published PR24 water base cost Stata do-file.

WAD_MSOAtoLAD_area: weighted average density calculated from MSOA data mapped to LAD level, using LAD area for the weights (code: BN4014), as reported in the published PR24 water base cost Stata do-file.

WAD_MSOAtoLAD_area2: quadratic (squared) term for the log of the weighted average density calculated from MSOA data mapped to LAD level, using LAD area for the weights, as reported in the published PR24 water base cost Stata do-file.

Brief comment on the models

We propose two sets of four models for consideration. The proposed models are identical to the PR19 models except for the density measure used of which there are two variations.

SRNWW1 to SRNWW4

In these four models we replace the Weighted Average Density (WAD) of PR19, which was based on density information at the local authority district (LAD) level, with a WAD that is based on density information at the Middle Layer Super Output Area (MSOA) level.

Ofwat developed the WAD at PR19 to replace the PR14 measure of density. The PR14 density variable measured average (unweighted) density across the entire water company area (or length of pipe). The PR14 measure was neutral to whether an area had a uniform mild density across or whether it had pockets of high and low densities, as long as the total number of customers per square mile was the same. In practice, infrastructure costs between these two types of areas are likely to be different (e.g. due to opportunities to build at scale where there are pockets of density and thereby benefit from economies of scale).

The PR19 WAD was a step in the right direction. It was an improvement on the PR14 density variable.

We consider that MSOA level information may allow a further improvement in the WAD. The MSOA is a smaller geographical unit than the LAD, without being ‘too small’ to capture economies of scale effect on treatment works (MSOAs are between 5,000 and 15,000 people). The MSOA better captures pockets of density and sparsity within a region hence may provide a more accurate information on regional density in water companies' service areas.

SRNWW5 to SRNWW8

In these four models we replace the WAD of PR19 with a WAD that is conceptually the same, but the information about the LAD characteristics (e.g. population, density, and in particular the proportion of the LAD that is attributed to a specific water company) is built up from MSOA level information. That is, there is an MSOA to LAD mapping, and subsequently a LAD to water company mapping.

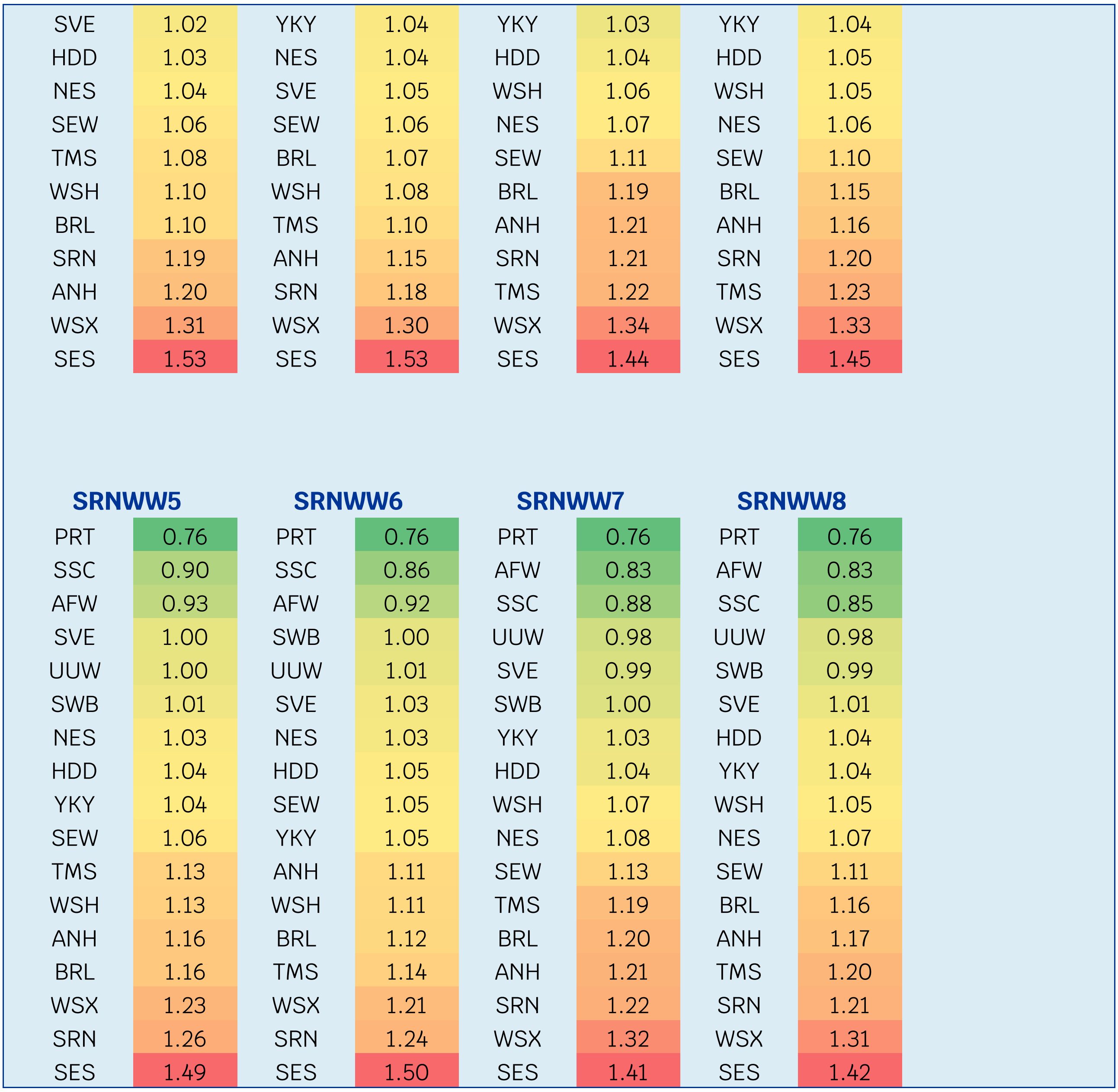
The main advantage of the MSOA to LAD mapping is that the attribution of LADs to water companies is more accurate where a water company straddles across several LADs. This makes the variable more accurate.

	SRNWW1	**SRNWW2**	**SRNWW3**	**SRNWW4**	**SRNWW5**	**SRNWW6**	**SRNWW7**	**SRNWW8**
Dependent variable	ln(botex plus wholesale water$_{it}$)	ln(botex plus wholesale water$_{it}$)	ln(botex plus wholesale water$_{it}$)	ln(botex plus wholesale water$_{it}$)	ln(botex plus wholesale water$_{it}$)	ln(botex plus wholesale water$_{it}$)	ln(botex plus wholesale water$_{it}$)	ln(botex plus wholesale water$_{it}$)
Properties (log)	1.052*** {0.000}	1.046*** {0.000}	1.105*** {0.000}	1.089*** {0.000}	1.072*** {0.000}	1.061*** {0.000}	1.110*** {0.000}	1.094*** {0.000}
Percentage of water treated in complexity levels 3-6 (%)	0.003*** {0.011}		0.002** {0.027}		0.003*** {0.002}		0.002** {0.029}	
Water treatment complexity index (log)		0.322** {0.030}		0.301* {0.053}		0.354** {0.016}		0.295* {0.058}
Booster per length (log)	0.509*** {0.003}	0.486*** {0.003}	0.518*** {0.003}	0.495*** {0.002}	0.457*** {0.008}	0.444*** {0.005}	0.519*** {0.003}	0.497*** {0.001}
WAD_MSOA_population (log)	-4.68*** {0.001}	-4.31*** {0.002}						
WAD_MSOA_population2 (log)	0.301*** {0.000}	0.276*** {0.001}						
WAD_MSOA_area (log)			-2.313** {0.021}	-2.035* {0.027}				
WAD_MSOA_area2 (log)			0.201** {0.020}	0.176** {0.025}				
WAD_MSOAtoLAD_population (log)					-1.85*** {0.000}	-1.65*** {0.001}		
WAD_MSOAtoLAD_population2 (log)					0.132*** {0.000}	0.117*** {0.000}		
WAD_MSOAtoLAD_area (log)							-2.56*** {0.009}	-2.28** {0.012}

WAD_MSOAtoLAD_area2 (log)							0.223*** {0.008}	0.198** {0.011}
Constant	10.300* {0.056}	8.674 {0.108}	-1.936 {0.413}	-2.821 {0.190}	-1.958 {0.206}	-2.795* {0.064}	-1.318 {0.572}	-2.208 {0.310}
Estimation method (OLS or RE)	RE	RE	RE	RE	RE	RE	RE	RE
N (sample size)	187	187	187	187	187	187	187	187
Model robustness tests								
R2 overall	0.96	0.97	0.96	0.96	0.97	0.97	0.96	0.96
RESET test	0.18	0.08	0.11	0.04	0.16	0.08	0.13	0.07
VIF (max)	506	527	307	318	205	206	321	334
Pooling / Chow test	0.99	0.97	0.99	0.98	0.94	0.86	0.99	0.97
Normality of model residuals	0.51	0.57	0.84	0.82	0.27	0.58	0.89	0.91
Heteroskedasticity of model residuals	0	0	0	0	0	0	0	0
Test of pooled OLS versus Random Effects (LM test)	0	0	0	0	0	0	0	0
Efficiency score distribution (min and max)	Max: 1.53 Min: 0.73 Gap: 0.79	Max: 1.53 Min: 0.74 Gap: 0.79	Max: 1.44 Min: 0.74 Gap: 0.70	Max: 1.45 Min: 0.74 Gap: 0.70	Max: 1.49 Min: 0.76 Gap: 0.73	Max: 1.50 Min: 0.76 Gap: 0.74	Max: 1.41 Min: 0.76 Gap: 0.64	Max: 1.42 Min: 0.76 Gap: 0.66
Sensitivity of estimated coefficients to removal of most and least efficient company	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**
Sensitivity of estimated coefficients to removal of first and last year of the sample	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**

Efficiency scores distribution

SRNWW1		**SRNWW2**		**SRNWW3**		**SRNWW4**	
PRT	0.73	PRT	0.74	PRT	0.74	PRT	0.74
SSC	0.92	SSC	0.88	AFW	0.84	AFW	0.84
AFW	0.93	AFW	0.92	SSC	0.89	SSC	0.86
SWB	0.99	SWB	0.98	UUW	0.98	SWB	0.98
UUW	1.02	UUW	1.02	SVE	0.99	UUW	0.99
YKY	1.02	HDD	1.04	SWB	1.00	SVE	1.01



SVE	1.02	YKY	1.04	YKY	1.03	YKY	1.04
HDD	1.03	NES	1.04	HDD	1.04	HDD	1.05
NES	1.04	SVE	1.05	WSH	1.06	WSH	1.05
SEW	1.06	SEW	1.06	NES	1.07	NES	1.06
TMS	1.08	BRL	1.07	SEW	1.11	SEW	1.10
WSH	1.10	WSH	1.08	BRL	1.19	BRL	1.15
BRL	1.10	TMS	1.10	ANH	1.21	ANH	1.16
SRN	1.19	ANH	1.15	SRN	1.21	SRN	1.20
ANH	1.20	SRN	1.18	TMS	1.22	TMS	1.23
WSX	1.31	WSX	1.30	WSX	1.34	WSX	1.33
SES	1.53	SES	1.53	SES	1.44	SES	1.45

SRNWW5		**SRNWW6**		**SRNWW7**		**SRNWW8**	
PRT	0.76	PRT	0.76	PRT	0.76	PRT	0.76
SSC	0.90	SSC	0.86	AFW	0.83	AFW	0.83
AFW	0.93	AFW	0.92	SSC	0.88	SSC	0.85
SVE	1.00	SWB	1.00	UUW	0.98	UUW	0.98
UUW	1.00	UUW	1.01	SVE	0.99	SWB	0.99
SWB	1.01	SVE	1.03	SWB	1.00	SVE	1.01
NES	1.03	NES	1.03	YKY	1.03	HDD	1.04
HDD	1.04	HDD	1.05	HDD	1.04	YKY	1.04
YKY	1.04	SEW	1.05	WSH	1.07	WSH	1.05
SEW	1.06	YKY	1.05	NES	1.08	NES	1.07
TMS	1.13	ANH	1.11	SEW	1.13	SEW	1.11
WSH	1.13	WSH	1.11	TMS	1.19	BRL	1.16
ANH	1.16	BRL	1.12	BRL	1.20	ANH	1.17
BRL	1.16	TMS	1.14	ANH	1.21	TMS	1.20
WSX	1.23	WSX	1.21	SRN	1.22	SRN	1.21
SRN	1.26	SRN	1.24	WSX	1.32	WSX	1.31
SES	1.49	SES	1.50	SES	1.41	SES	1.42

Proposed model for Economies of scale in wastewater treatment

Econometric model formula:

1.
$\ln(real\ botex\ plus\ sewage\ treatment_{it}) = \beta_0 + \beta_1 \ln load_{it} + \beta_2\, pctnh3below3mg_{it} + \beta_3\, WAWs_{it} + \varepsilon_{it}$

2.
$\ln(real\ botex\ bioresources\ plus_{it}) = \beta_0 + \beta_1 \ln load_{it} + \beta_2\, pctnh3below3mg_{it} + \beta_3\, WAWs_{it} + \varepsilon_{it}$

Description of the dependent variable

The dependent variables are: lnrealbotexplusswt and lnrealbotexplusbrp, which are exactly the same as the dependent variables used at PR19 for the sewage treatment and bioresources models respectively.

We acknowledge that Ofwat has indicated that it will model bioresources separately at PR24. We therefore include a bioresources version of the model, firstly because it demonstrates the merit of the new variable and secondly to show that the bioresources model above could be used if it were to remain relevant at PR24.

Description of the explanatory variables

Sewage Treatment 1:

- Ln(load): natural logarithm of load (STWDP125_21), as reported in the published wholesale dataset.
- pctnh3below3mg: the percentage of ammonia consent (from summing STWDA121 & STWDA122_21 then dividing by load) , as reported in the published wholesale dataset.
- WAWS: the weighted average of STW sizes, in logs. The derivation of the variable is explained below.

Bioresources plus:

- Ln(load): as above.

- pctnh3below3mg: as above.
- WAWS: as above.

Context

The rationale for including economies of scale variables in sewage treatment models was not disputed at PR19: there is both economic rationale, engineering rationale and empirical evidence that unit cost tends to decrease with works' scale continuously across all scales (albeit in decreasing rates).

What was disputed is the specific choice of variables. The variables selected assume a 'step change' in unit cost between certain size bands (e.g., between band 3 and band 4) and do not give sufficient representation to the very wide range of scales across works in the sector, especially across the largest band where the vast majority of treatment activity occurs. As a consequence, the models do not fully capture the relationship between works' scale and cost.

This issue has been raised in the past by Severn Trent Water[1] and South West Water[2] in response to Ofwat's econometric consultation at PR19.

More recently, Anglian Water[3] has raised the particular concern that Ofwat's PR19 models do not adequately capture the impact of economies of scale of very large works on treatment costs. As seen in Table 1, the 'above band 5' category is 'open-ended'. This category includes a very wide range of sizes and, importantly, nearly 80% of sewage load in the sector. Consequently, Ofwat's variable 'the proportion of load treated at above band 5', lumps together a wide range of large works and, as such, assumes that economies of scale (i.e., average costs) are the same across all works in this band.

Anglian proposed to disaggregate the 'above band 5' category to five new bands, 6 to 10. It proposed using the proportion of load received at band 8 or above as a cost driver to replace the PR19 variable 'the proportion of load treated at above band 5'. A concern with this approach is the need to define new bands and cut-offs: how many additional bands to use? where to set the cut-offs between bands? where to arbitrarily set the cut-off for the new variable?

Anglian presented clear evidence that economies of scale in wastewater treatment across the whole range of bands, albeit at diminishing rate (which is the property of a normal cost function) – see the figure below.

---

[1] SVT-consultation-response.pdf (ofwat.gov.uk), paragraph 84(2), page 26.

[2] SWT-consultation-response.pdf (ofwat.gov.uk), Section 5.2, page 17.

[3] Cost modelling advantage of Band 6 WRC disaggregation, Anglian Water, presented at Ofwat's cost assessment working group, November 21.

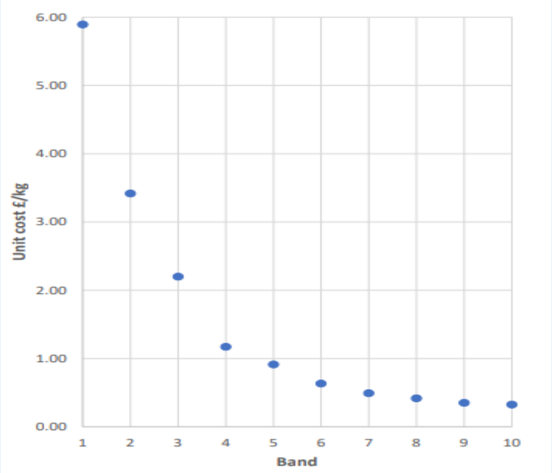


Figure 1: Weighted average unit cost by size band (2020-21)

Source: Ofwat, Cost Assessment Working Group slides, 21 November.

Despite the diminishing rates of economies of scale at bands 6-10, the fact that almost 80% of load is received at this range means that small variation in STW sizes across companies at this range can have a material impact on total treatment costs. It is therefore important to capture this variation as an effective explanatory variable.

Derivation of WAWS

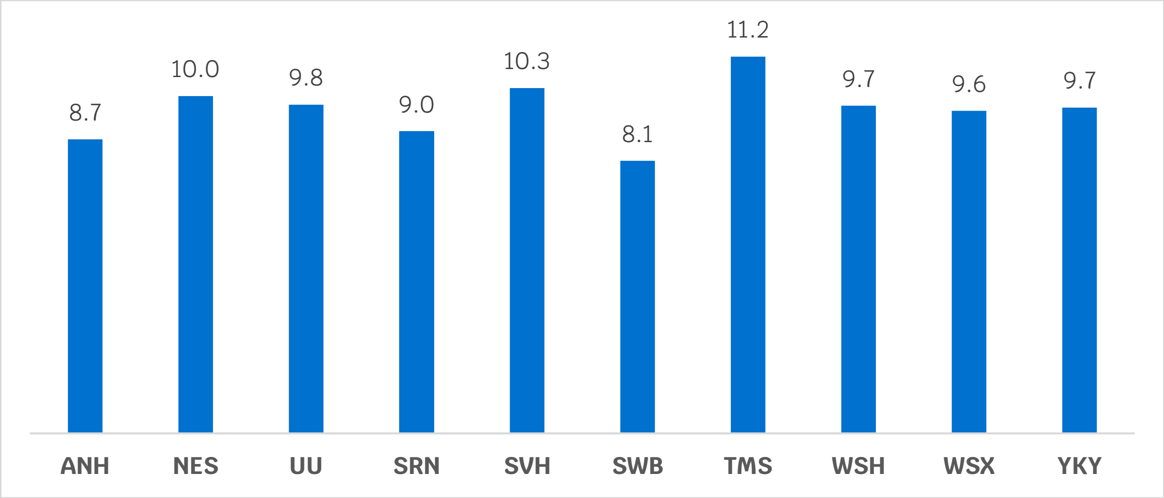
We have developed an alternative measure to capture the size of STWs at each company in a more holistic way, by using all the data currently available through the APRs and without needing to arbitrarily define any thresholds. The measure is the ‘weighted average works size’ (WAWS). WAWS is the weighted average of STW sizes, in logs, where for works at bands 1-5 the size is the midpoint of the band. For example, for band 2 it is the midpoint between 15 and 30, i.e. 22.5. For works at above band 5 the size is the actual size of each individual STW as this information has been regularly collected through the APRs. The weights are the proportion of kg BOD5 load received at each size.

The formula for the WAWS, in kg BOD5/day, is:

$$\begin{aligned} WAWS_i = {} & 7.5 * (\%\, load\ received\ at\ band\ 1\ for\ company\ i) + 22.5 \\ & * (\%\, load\ received\ at\ band\ 2\ for\ company\ i) + 75 \\ & * (\%\, load\ received\ at\ band\ 3\ for\ company\ i) + 360 \\ & * (\%\, load\ received\ at\ band\ 4\ for\ company\ i) + 1050 \\ & * (\%\, load\ received\ at\ band\ 5\ for\ company\ i) \\ & + \sum_j (size\ of\ STW_j) * (\%\, load\ received\ at\ STWj\ of\ company\ i) \end{aligned}$$

In this formula, *i* denotes a company, j denotes a STW above band 5. The numbers 7.5, 22.5 etc. are the mid-points of the respective band, measured in kg BOD5/day. We then take the natural log of $WAWS_i$. A representation of the variable is given in Figure 2.

Figure: The weighted average works size (WAWS) variable (in log based on 2020-21 data)



The WAWS has the benefit that it uses all the information available in the APRs to construct a holistic measure of STWs' average size at each company without any additional cost to companies to collect the information. The WAWS variable gives manifestation to each and every band (or, for works above band 5, between each and every works); it does not require making unnecessary assumptions on which bands to group together, which in turn arbitrarily imposes a rigid relationship between works' size and cost (for example grouping bands 1-3 imposes the assumption that all works in these bands have the same economies of scale and all works above band 3 have the same economies of scale).

The weights, measured by the proportion of total load at each size band, ensure that the variables provide the best representation of the company's economies of scale in sewage treatment.

A potential alternative to the WAWS is a variable that applies a log to each sewage treatment works' size (or to each mid-band size for bands 1-5) and uses similar weights as the WAWS to calculate a weighted average of logged sizes. This alternative variable works equally well as the WAWS in the models.

In both variants, the log captures the diminishing rate of economies of scale as size increases.

Brief comment on the models

We tested the WAWS in wastewater models that include sewage treatment costs. That is, in the bioresources plus and the wastewater treatment models.

The table below provides estimation results. The results are based on a ‘random effects’ estimation using panel data from 2011-12 to 2020-21. The table presents the PR19 models (updated with new data) and an alternative model which includes our proposed variable, the WAWS.[4]

The WAWS is highly statistically significant, with a plausible magnitude and the expected sign. The R-squared is showing a significant improvement compared to the PR19 models. It rises from 0.88/0.87 to 0.92 and 0.92 to 0.95. The models with the WAWS variable pass the RESET test more robustly than the PR19 models, and the range of efficiency scores narrows from 0.61/0.62 to 0.32 and 0.48/0.45 to 0.31.

By making a fuller use of information reported in APRs, we consider that the alternative variable better reflects the range of STWs sizes in the sector, and therefore better captures the impact of economies of scale on cost. The statistical evidence strongly supports the use of our proposed alternative variable.

From an engineering perspective there are well understood economies of scale as treatment works increase in size. Given that 82-83% of the industry’s load is treated in the largest Band 6 works group (for all years from 2012-2021), this variable separates out the effect in a more granular fashion by using existing information of the size of each large works, The model coefficients are consistent with this logic.

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Dependent variable	lnrealbotexplusswt	lnrealbotexplusbrp			
Load (log)	0.758***	0.832***			
Weighted average works size (WAWS)	-0.224***	-0.211***			
Load with ammonia consent below 3mg/l (%)	0.006***	0.005***			
Constant	-2.800***	-3.459***			
Estimation method (OLS or RE)	RE	RE			
N (sample size)	110	110			
Model robustness tests					
R2 adjusted	0.91	0.95			
RESET test	0.50	0.25			

---

[4] The WAWS variable in our analysis is based on data from the year 2020-21. The same value was used for all years of the panel data. We tested the WAWS with the three years of data (218-19 to 202-21) and found the variable to be equally robust and significant. The APRs include information to construct a complete time series of the variable for use at PR24.

VIF (max)	4.35[#]	4.35[#]			
Pooling / Chow test	1	0.81			
Normality of model residuals	0.06	0.00			
Heteroskedasticity of model residuals	0.96	0.53			
Test of pooled OLS versus Random Effects (LM test)	0	0			
Efficiency score distribution (min and max)	0.36	0.32			
Sensitivity of estimated coefficients to removal of most and least efficient company	G	G			
Sensitivity of estimated coefficients to removal of first and last year of the sample	G	G			

Efficiency scores distribution

Company	SWT1	SWT2
ANH	0.93	0.96
NES	1.06	0.94
NWT	1.04	0.94
SRN	1.26	1.10

SVT	0.96	0.97
SWB	0.90	0.93
TMS	0.95	0.97
WSH	1.20	1.25
WSX	1.02	1.01
YKY	1.05	1.16
Range	0.36	0.32

Proposed model for sewage treatment complexity variable

Econometric model formula:

SRNSTW1. $\ln(\text{WWW botex plus sewage treatment}_{it}) = \alpha + \beta_1 \ln(\text{load}_{it}) + \beta_2 (\text{percentage of load treated in STWs bands 1-3}_{it}) + \beta_3 \ln(\text{percentage of load with phosphorus below 0.5mg/l + BOD5 below 7mg/l + ammonia below 1mg/l}_{it}) + \varepsilon_{it}$

SRNSTW2. $\ln(\text{WWW botex plus sewage treatment}_{it}) = \alpha + \beta_1 \ln(\text{load}_{it}) + \beta_2 (\text{percentage of load treated in STWs band 6}_{it}) + \beta_3 \ln(\text{percentage of load with phosphorus below 0.5mg/l + BOD5 below 7mg/l + ammonia below 1mg/l}_{it}) + \varepsilon_{it}$

SRNSTW3. $\ln(\text{WWW botex plus sewage treatment}_{it}) = \alpha + \beta_1 \ln(\text{load}_{it}) + \beta_2 (\text{percentage of load treated in STWs bands 1-3}_{it}) + \beta_3 \ln(\text{percentage of load with phosphorus below 0.5mg/l + BOD5 below 7mg/l + ammonia between 1mg/l and 3mg/l}_{it}) + \varepsilon_{it}$

SRNSTW4. $\ln(\text{WWW botex plus sewage treatment}_{it}) = \alpha + \beta_1 \ln(\text{load}_{it}) + \beta_2 (\text{percentage of load treated in STWs band 6}_{it}) + \beta_3 \ln(\text{percentage of load with phosphorus below 0.5mg/l + BOD5 below 7mg/l + ammonia between 1mg/l and 3mg/l}_{it}) + \varepsilon_{it}$

Description of the dependent variable

Wholesale wastewater botex plus sewage treatment (code: botexplusswt). That is, botexswt + reduce flooding risk for properties capex (S3023ST), as reported in the published PR24 wastewater base cost Stata do-file.

Description of the explanatory variables

load: load using Phosphorus total load (code: STWDP125_21), as reported in the published wholesale wastewater dataset.

percentage of load treated in STWs bands 1-3 (code: pctbands13), as reported in the published wholesale wastewater dataset.

percentage of load treated in STWs band 6 (code: pctbands6), as reported in the published wholesale wastewater dataset.

percentage of load with phosphorus below 0.5mg/l + BOD5 below 7mg/l + ammonia below 1mg/l. This is a sum of the load received at the strictest consent levels for phosphorus (below 0.5mg/l), BOD5 (below 7mg/l) and ammonia (below 1mg/l) as a percentage of the total load. That is, (STWDP121_21 + STWDB121_21 + STWDA121) / load* 100

percentage of load with phosphorus below 0.5mg/l + BOD5 below 7mg/l + ammonia between 1mg/l and 3mg/l. This is a sum of the load received at strict consent levels for phosphorus (below 0.5mg/l), BOD5 (below 7mg/l) and ammonia (between 1mg/l and 3mg/l) as a percentage of the total load. That is, (STWDP121_21 + STWDB121_21 + STWDA122_21) / load* 100

codes: STWDP121_21, STWDB121_21, STWDA121 and STWDA122_21, as reported in the published wholesale wastewater dataset.

Brief comment on the models

Treatment complexity is a key driver of sewage treatment costs – the tighter the consent level, the more costly it is to comply with it due to the use of more demanding treatment processes.

The PR19 treatment complexity variable accounted only for consent levels related to ammonia. However, ammonia is not the only pollutant that needs to be removed at cost, phosphorous and BOD also require removal. Removal of phosphorous is becoming more and more material as part of base costs due to tightening P consents across the sector and the additional expenditure required to comply with it.

To this end, we propose models that replace the PR19 treatment complexity variable with a single ‘composite’ variable that accounts for ammonia, phosphorous and BOD.

Our proposed treatment complexity variable combines all three chemicals at the strictest consent levels reported in the APRs. This variable improves the overall fit of the sewage treatment models and reduces the range of the efficiency scores, when compared to models that use the PR19 variable that only accounts for ammonia consent below 3mg.

Not only is this more holistic variable sound from an engineering and statistical perspective, but it is also more appropriate for the purpose of setting future allowances given the dynamism in level of consents across the sector, and that these consents are largely an exogenous requirement on water companies.

	SRNSTW1	SRNSTW2	SRNSTW3	SRNSTW4
Dependent variable	ln(botex plus sewage treatment$_{it}$)	ln(botex plus sewage treatment$_{it}$)	ln(botex plus sewage treatment$_{it}$)	ln(botex plus sewage treatment$_{it}$)
Load (log)	0.790*** {0.000}	0.791*** {0.000}	0.664*** {0.000}	0.669*** {0.000}
Load treated in size bands 1-3 (%)	0.031 {0.126}		0.03 {0.195}	
Load treated in size band 6 (%)		-0.009 {0.126}		-0.009 {0.091}
Load with phosphorus <0.5mg/l + BOD5 <7mg/l + ammonia <1mg/l (%)	0.009** {0.039}	0.010** {0.039}		
Load with phosphorus <0.5mg/l + BOD5 <7mg/l + ammonia >1mg/l & <3mg/l (%)			0.006*** {0.000}	0.006** {0.000}
Constant	-5.381*** {0.000}	-4.572*** {0.000}	-3.875*** {0.007}	-3.087*** {0.001}
Estimation method (OLS or RE)	RE	RE	RE	RE
N (sample size)	110	110	110	110
Model robustness tests				
R2 overall	0.87	0.87	0.85	0.85
RESET test	0.00	0.01	0.03	0.07
VIF (max)	2.84	2.20	4.95	4.05
Pooling / Chow test	0.77	0.93	1.00	1.00
Normality of model residuals	0.11	0.22	0.03	0.07
Heteroskedasticity of model residuals	0.31	0.67	0.42	0.89
Test of pooled OLS versus Random Effects (LM test)	0	0	0	0
Efficiency score distribution (min and max)	Max: 1.42 Min: 0.85 Gap: 0.58	Max: 1.42 Min: 0.88 Gap: 0.55	Max: 1.45 Min: 0.81 Gap: 0.64	Max: 1.45 Min: 0.83 Gap: 0.62
Sensitivity of estimated coefficients to removal of most and least efficient company	G	G	G	G
Sensitivity of estimated coefficients to removal of first and last year of the sample	G	G	G	G

Efficiency scores distribution

SRNSTW1		SRNSTW2		SRNSTW3		SRNSTW4	
TMS	0.85	TMS	0.88	TMS	0.81	TMS	0.83
NES	0.95	WSX	0.91	SVE	0.89	SVE	0.88
WSX	0.95	SVE	0.95	SWB	0.94	WSX	0.91
SVE	0.95	NES	0.97	NES	0.94	SWB	0.97
SWB	0.98	ANH	0.99	WSX	0.95	NES	0.98
WSH	1.01	SWB	1.02	WSH	1.06	ANH	1.04
UUW	1.06	WSH	1.04	YKY	1.09	YKY	1.05
ANH	1.06	UUW	1.08	ANH	1.12	WSH	1.10
YKY	1.20	YKY	1.17	UUW	1.23	UUW	1.27
SRN	1.42	SRN	1.42	SRN	1.45	SRN	1.45

Proposed model for coastal variable

Econometric model formula:

SRNSTW5. $\ln(\text{WWW botex plus sewage treatment}_{it}) = \alpha + \beta_1 \ln(\text{load}_{it}) + \beta_2$ (percentage of load treated in STWs bands 1-3$_{it}$) + β_3 (percentage of load with ammonia consent below 3mg/l$_{it}$) + β_4 (percentage of coastal population) + ε_{it}

SRNSTW6. $\ln(\text{WWW botex plus sewage treatment}_{it}) = \alpha + \beta_1 \ln(\text{load}_{it}) + \beta_2$ (percentage of load treated in STWs band 6$_{it}$) + β_3 (percentage of load with ammonia consent below 3mg/l$_{it}$) + β_4 (percentage of coastal population) + ε_{it}

Description of the dependent variable

Wholesale wastewater botex plus sewage treatment (code: botexplusswt). That is, botexswt + reduce flooding risk for properties capex (S3023ST), as reported in the published PR24 wastewater base cost Stata do-file.

Description of the explanatory variables

load: load using Phosphorus total load (code: STWDP125_21), as reported in the published wholesale wastewater dataset.

percentage of load treated in STWs bands 1-3 (code: pctbands13), as reported in the published wholesale wastewater dataset.

percentage of load treated in STWs band 6 (code: pctbands6), as reported in the published wholesale wastewater dataset.

percentage of load with ammonia consent below 3mg/l (code: pctnh3below3mg), as reported in the published wholesale wastewater dataset.

percentage of coastal population: the proportion of the population within a company service area that is coastal, as reported by the ONS. The datasets are from the ONS, namely coastal towns [Coastal towns in England and Wales - Office for National Statistics (ons.gov.uk)] and coastal cities [available on request from ONS: Subnational@ons.gov.uk]. We used 100% of the population allocated to a single LAD, as per advice from ONS and used the mapping provided by ONS to our query. Please also find attached how we derived the variable in the file called "Coastal variable.xlsb".

Brief comment on the models

Serving coastal population has unique challenges, which present specific cost pressures to wastewater treatment. Below we set out factors that underlie these cost pressures.

To capture these factors, we propose a new variable to Ofwat's PR19 wastewater treatment models – the proportion of 'coastal' population in a company area, based on ONS data.

The variable is exogenous, statistically significant with the right sign and plausible magnitude. The variable improves models' quality and performance; it has a strong engineering rationale.

The variable has a perceptible, significant, and logical influence on wastewater treatment cost models.

Engineering rationale

There are several factors that put cost pressures on coastal treatment works. We outline these factors below.

1. Requirements on effluent quality

Some restrictions on wastewater discharge are common in sewage treatment works (STWs) that discharge to inland waters, but not in STWs that discharge to seawaters. This is the case with restrictions on the discharge of ammonia and phosphorous. Other restrictions are more common in STWs that discharge to coastal waters, particularly those close to bathing or shellfish waters. This is the case with UV treatment (or other forms of disinfection) or a total nitrogen consent. Evidence for this can be found in the APRs (table 7B) where UV disinfection requirements are found only for coastal wastewater companies, but not for inland companies such as Thames or Severn Trent. UV disinfection imparts additional tertiary treatment cost. Total nitrogen consent imparts additional costs (as it requires internal recirculation at additional pumping costs). While on their own, these costs may not be material enough to render UV or N-consent as statistically significant cost drivers, it is important to recognise that they provide systematic additional cost due to factors beyond management control for coastal companies. Ofwat's PR19 models capture only requirements on discharges to inland waters, which exacerbates the issue and creates a bias.

2. Saline environment

Enhanced corrosion from saline water and salt spray drives higher maintenance costs. These costs relate to higher specification valves and mechanical parts to cope with the corrosive environment, more frequent replacement of corroded assets and painting rusting structures. We can provide case studies to evidence the above. Based on 2020-21 data, our large coastal works on average incur 40% higher repair costs than inland ones (per unit of load).

Saline water contains higher levels of sulphate than non-saline water, leading to higher risk of hydrogen sulphide creation during wastewater treatment. In a poorly ventilated space, this will result in rapid corrosion of not only mechanical, electrical and ICA equipment, but also concrete. To combat this, higher grade materials with better corrosion resistance have to be used, and enhanced ventilation and odour control is needed.

Coastal works require increased chemical dosing to combat the production of hydrogen sulphide. For 2020-21 data, chemical costs at our large coastal works were 71% higher, on a per unit of load basis, than at inland works.

3. Space constraints

Much of the Southern Water coastline is heavily populated, with little sparsely occupied land around the population centres, particularly as the urban areas are constrained by the sea on at least one side. This leads to two general STW designs – either being located within urban areas or to move the STW inland and pump waste uphill and a significant distance (see “Double pumping” below). By contrast many inland works are located downstream of a conurbation at a sufficient distance to avoid odour issues and allowing gravity sewers to deliver the wastewater. In constrained coastal locations we don’t have that option and Local Authority planning regulations require the works in urban areas to be covered to prevent odour issues affecting the nearby population. Space constraints therefore lead to additional costs related to odour restrictions, retrofitting works on constrained sites, maintaining covered sites and dealing with additional corrosion from hydrogen sulphide. (The latter is due to long rising mains which often feed coastal STWs. Lack of air entrainment in rising mains results in septic sewage which produces H2S, with nowhere for the gas to disperse in covered works.)

4. Double pumping

Traditionally coastal treatment works only had preliminary or primary treatment before being discharged to sea. In the 1990s, secondary treatment was required before discharge, which required much more space. This was problematic for many of our coastal sites which had a small footprint and were situated in coastal urban areas. One solution (discussed above) was to retrofit a very compact treatment works on the original site and cover or bury it to comply with odour restrictions. Our treatment works at Eastbourne, which is underground at the end of the promenade is a good example.

Another solution was to pump the flows inland to a new STW site, and then pump back to the original seafront location to discharge using the original sea outfalls. Examples include, Weatherlees Hill (serving Margate, Broadstairs, Ramsgate, Deal, Sandwich); Ford (serving Bognor Regis, Littlehampton); Budds Farm (serving Portsmouth); Broomfield Bank (serving

Dover, Folkestone); Peacehaven (serving Brighton and Hove); and West Hythe. These works treat 25% of flows.
Double pumping all flows adds significant power costs compared to conventional treatment works. Sampling 194 of our STWs we find that coastal STWs have power cost per load that are 70% higher than inland STWs.

5. Peakiness (i.e., large variation around average load)

Many coastal areas experience extreme summer peaks due to tourism. STWs must be sized based on peak load (structure and treatment assets). Ofwat's models use total load as a cost driver, however, this variable does not capture the effect of peakiness: for two STWs with identical total annual load, the one that has higher peak would be larger, with higher maintenance and operation costs both at peak and off-peak periods (when small load is treated with an over-sized works).

6. Outfalls

STWs discharging to an inland river tend to have a gravity outfall at the back of the STW requiring no mechanical or electrical operation. STWs that discharge to seawater tend to have multiple and longer piped outfalls compared to inland works. Sea outfalls are usually over 1km long and incur higher maintenance costs including offshore navigation maintenance requirements. They also require pumping of the full STW load during both normal and storm conditions along with requisite backup pumps. For example, our long sea outfall serving Portsmouth's STW is 3.5km and requires pumping at a maximum rate of 311 Ml/d.

7. Spill frequency

STWs that discharge to seawaters have stricter spill frequency constraints. As a result, more storm tank, storm screening and storm pumping capacity is required with additional pumping to store and then treat the extra flow, and additional maintenance costs.

Spill design frequency criteria for bathing waters is three per bathing season and for shellfish water 10 per year. Inland STWs do not have such stringent constraints.
The following data extracted from reports by the Environment Agency1,2 identifies spill frequency investigation triggers.

Receiving water body	Spills/Year or Bathing Season
Fresh waters	
1 year of EDM	60
2 years of EDM	50
3 or more years of EDM	40
Sea waters	

---

[1] Storm Overflows Assessment Framework, Environment Agency, June 2018.

[2] Water companies: environmental permits for storm overflows and emergency overflows, Environment Agency, September 2018, link.

Shellfish Water	10
Bathing Water	3

8. Resilience costs

Coastal works have increased risk of sea rise and wave/tidal action, which require specific design specifications. Additional energy resilience is needed for coastal STWs given they are often at the end of the electricity distribution network with limited contingency.

Econometric performance

From an engineering perspective we expect the coastal effect to be particularly relevant for wastewater treatment costs. The models we propose here are for wastewater treatment. They are similar to Ofwat's PR19 models with the addition of the coastal variable to the specification. The results are based on a 'random effects' estimation using panel data from 2011-12 to 2021-22.

The coastal variable is highly statistically significant, with a plausible magnitude and the expected sign. The R-squared improves compared to the PR19 models moving from 0.878/0.88 to 0.90/0.92, and the range of efficiency scores narrows from 0.61/0.62 to 0.41/0.34. The impact on the other coefficients in the model is minimal, except for that of load, which significantly increases. The new coefficient estimate (circa 1) brings it more in line with the coefficient estimate of scale drivers in other water and wastewater models.

We tested the sensitivity of the models by excluding one year and company at a time. The coastal variable remains robust in both treatment models. The RESET test[3] becomes marginally insignificant with the new variable. We do not consider this as a reason to reject the new variable given its overall strengths. At PR19 Ofwat said "[a] failure of the reset test should prompt a search for a more flexible specification, but need not in itself be grounds for dismissing a model"[4], and in fact put forward sewage treatment models that fail the RESET test in its 2018 econometric consultation. Further, our sensitivity analysis found the RESET test passed at the 5% level in most model specifications. In the modelling guidance, Ofwat state "failure of the RESET test on its own may not be a valid justification to dismiss a model. This is particularly the case if it is considered that the model offers useful information from an economic or engineering perspective. The higher the p-value, the more confident we are that the functional form is adequate." The p-values being higher than previous model specifications alongside the unique engineering circumstances coastal works face that aren't included in other explanatory variables, supports the inclusion of this variable.

---

[3] The RESET test detects misspecification error (e.g. an omitted variable or the existence of non-linearities).

[4] Cost assessment for PR19: a consultation on econometric cost modelling, Ofwat, March 2018, page 11.

	SRNSTW5	SRNSTW6			
Dependent variable	ln(botex plus sewage treatment$_{it}$)	ln(botex plus sewage treatment$_{it}$)			
Load (log)	0.890*** {0.000}	0.956*** {0.000}			
Load treated in size bands 1-3 (%)	0.036** {0.019}				
Load treated in size band 6 (%)		-0.013*** {0.000}			
Load with ammonia consent below 3mg/l (%)	0.006*** {0.000}	0.006*** {0.000}			
Coastline population (%)	0.011*** {0.000}	0.014*** {0.000}			
Constant	-6.969*** {0.000}	-6.711*** {0.000}			
Estimation method (OLS or RE)	RE	RE			
N (sample size)	110	110			
Model robustness tests					
R2 adjusted	0.90	0.92			
RESET test	0.00	0.04			
VIF (max)	7.10	6.41			
Pooling / Chow test	0.98	0.98			
Normality of model residuals	0.68	0.49			
Heteroskedasticity of model residuals	0.06	0.16			
Test of pooled OLS versus Random Effects (LM test)	0	0			
Efficiency score distribution (min and max)	Max: 1.23 Min: 0.82 Gap: 0.41	Max: 1.20 Min: 0.86 Gap: 0.34			
Sensitivity of estimated coefficients to removal of most and least efficient company	**G**	**G**			
Sensitivity of estimated coefficients to removal of first and last year of the sample	**G**	**G**			

Efficiency scores distribution

SRNSTW5		SRNSTW6	
TMS	0.82	TMS	0.86
SWB	0.92	SWB	0.95

SVH	1.00	WSX	1.00
WSH	1.01	SVH	1.01
NES	1.03	ANH	1.03
NWT	1.03	NWT	1.05
WSX	1.03	WSH	1.05
ANH	1.15	NES	1.11
SRN	1.20	SRN	1.15
YKY	1.23	YKY	1.20