Appendix 8.2.A – Wholesale cost models

Wessex Water

September 2018

Wessex Water
YTL GROUP

Business plan section		Supporting document	
	Board vision and executive summary		
1	Engaging customers		
2	Addressing affordability and vulnerability		
3	Delivering outcomes for customers		
4	Securing long term resilience		
5	Markets & innovation: wholesale		
6	Markets & innovation: open systems & DPC		
7	Markets & innovation: retail		
8	Securing cost efficiency	8.1	Input cost and frontier shift assumptions
		8.2	**Wholesale cost modelling and the calculation of catch-up**
		8.3	Residential retail expenditure
		8.4	Cost adjustment claims covering letter
		8.5	Claim WSX01 summary – North Bristol sewerage strategy
		8.6	Claim WSX02 summary – Sewage treatment works capacity programme
		8.7	Claim WSX03 summary – Number of non-infrastructure water supply assets
		8.8	Claim WSX04 summary – Reducing leakage by a further 15%
		8.9	Claim WSX05 summary – Flooding programme
		8.10	Claim WSX06 summary – Pollution reduction strategy
		8.11	Assessing the costs of our enhancement programme
9	Aligning risk and return		
10	Financeability		
11	Accounting for past delivery		
12	Securing trust, confidence and assurance		
13	Data tables and supporting commentaries		

Template for submission of econometric models for consultation
Bioresources models

Econometric model formula:

1. $\ln(Expenditure) = \propto_0 + \propto_1 \ln(Sludge\ Produced) + \propto_2\ (Ofwat\ measure\ of\ highly\ dense\ areas) + time\ dummy\ variables$

2. $\ln(Smoothed\ expenditure) = \propto_0 + \propto_1 \ln(Sludge\ Produced) + \propto_2\ (Ofwat\ measure\ of\ highly\ dense\ areas) + time\ dummy\ variables$

3. $\ln\left(\frac{Expenditure}{sludge\ produced}\right) = \propto_0 + \propto_1\ (Ofwat\ measure\ of\ highly\ dense\ areas) + time\ dummy\ variables$

4. $\ln\left(\frac{Smoothed\ expenditure}{Sludge\ produced}\right) = \propto_0 + \propto_1\ (Ofwat\ measure\ of\ highly\ dense\ areas) + time\ dummy\ variables$

Description of dependent variable

Model 1 Bioresources botex := Opex + capital Maintenance – third party costs – local authority rates – EA charges

Model 2 Bioresources botex := Opex + IRE + average MNI over period – third party costs – local authority rates – EA charges

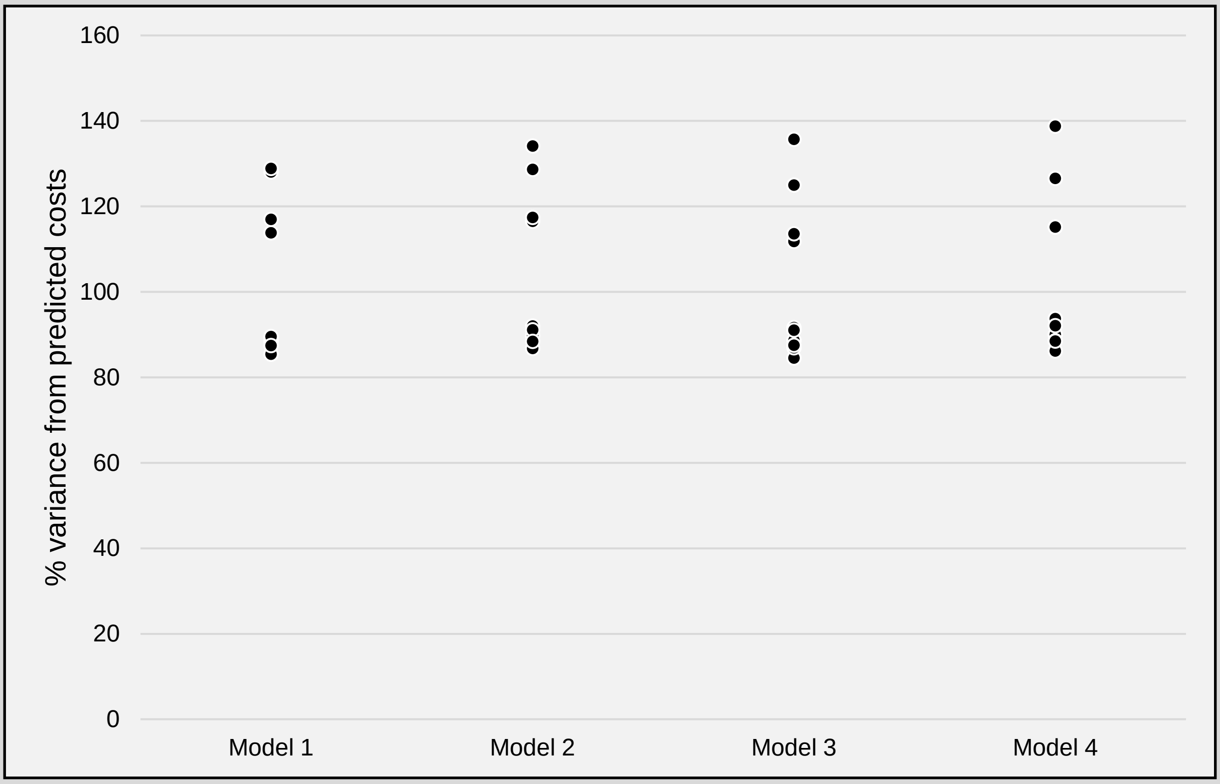
Brief comment on models

Simple, exogenous bioresources models. No endogenous variables were included to aid in setting a level playing field for market opening. Limited independent observations limits number of variables we could include. Please see attached note for more information.

Ofwat measure of highly dense areas := the proportion of the companies area of service with over 6000 pop. Density as defined in constructed dataset on the Ofwat SharePoint.

	Model 1	**Model 2**		
Dependent variable	Aggregate Bioresources Botex	Smoothed Bioresources Botex	Unit Bioresources Botex	Smoothed Bioresources Botex
Sludge Produced	0.9683*** (0.000)	0.9521*** (0.000)		
Ofwat measure of highly dense areas	-0.5290 (0.228)	-0.4738 (0.256)	-0.4105 (0.195)	-0.3839 (0.237)
Constant	-0.6598 (0.386)	-0.7359 (0.269)	-8.7601*** (0.000)	-8.9054*** (0.000)
R2 adjusted	0.8152	0.8899	0.0769	0.1399
VIF (max)	1.67	1.67	1.67	1.67
Reset test	0.000	0.000	0.6374	0.6615
Estimation method (eg OLS or RE)	OLS	OLS	OLS	OLS
N (sample size)	60	60	60	60

Chart


160
140
120
100
80
60
40
20
0
% variance from predicted costs
Model 1
Model 2
Model 3
Model 4

Wessex Water submission of econometric models for consultation
Residential retail

Econometric model formula:

A1. ln(*total retail operating costs*$_{it}$) = β_0 + β_1 ln(*single service customers*$_{it}$) + β_2 ln(*dual service customers*$_{it}$) + β_3 *flats*$_{it}$ + β_4 *IMD income*$_{it}$ + β_5 ln(*average wholesale bill*$_{it}$) + ε_{it}

A2. ln(*bad debt related operating costs*$_{it}$) = β_0 + β_1 ln(*single service customers*$_{it}$) + β_2 ln(*dual service customers*$_{it}$) + β_3 *IMD income*$_{it}$ + β_4 ln(*average wholesale bill*$_{it}$) + β_5 *internal migration*$_{it}$ + ε_{it}

A3. ln(non-*bad debt related operating costs*$_{it}$) = β_0 + β_1 ln(*single service customers*$_{it}$) + β_2 ln(*dual service customers*$_{it}$) + β_3 *metered households*$_{it}$ + β_4 *metered household density*$_{it}$ + β_5 ln(peak traffic speed$_{it}$) + ε_{it}

A4. ln(*total retail operating costs*$_{it}$) = β_0 + β_1 ln(*single service customers*$_{it}$) + β_2 ln(*dual service customers*$_{it}$) + β_3 *metered households*$_{it}$ + β_4 *metered household density*$_{it}$ + β_5 *flats*$_{it}$ + β_6 ln(*peak traffic speed*$_{it}$) + β_7 *IMD income*$_{it}$ + β_8 ln(*average wholesale bill*$_{it}$) + ε_{it}

A5. ln(*total retail operating costs*$_{it}$) = β_0 + β_1 ln(*single service customers*$_{it}$) + β_2 ln(*dual service customers*$_{it}$) + β_3 *IMD income*$_{it}$ + β_4 *property repossessions*$_{it}$ + β_5 ln(*average wholesale bill*$_{it}$) + u_i + v_{it}

A6. ln(*bad debt related operating costs*$_{it}$) = β_0 + β_1 ln(*single service customers*$_{it}$) + β_2 ln(*dual service customers*$_{it}$) + β_3 *IMD income*$_{it}$ + β_4 ln(*average wholesale bill*$_{it}$) + u_i + v_{it}

A7. ln(non-*bad debt related operating costs*$_{it}$) = β_0 + β_1 ln(*single service customers*$_{it}$) + β_2 ln(*dual service customers*$_{it}$) + β_3 *metered households*$_{it}$ + β_4 ln(peak traffic speed$_{it}$) + β_5 *time trend*$_t$ + u_i + v_{it}

A8. ln(*total retail operating costs*$_{it}$) = β_0 + β_1 ln(*single service customers*$_{it}$) + β_2 ln(*dual service customers*$_{it}$) + β_3 *metered households*$_{it}$ + β_4 *flats*$_{it}$ + β_5 *IMD income*$_{it}$ + β_6 *property repossessions*$_{it}$ + β_7 ln(*average wholesale bill*$_{it}$) + u_i + v_{it}

B1. ln(*total retail operating costs*$_{it}$) = β_0 + β_1 ln(*total customers*$_{it}$) + β_2 *IMD income*$_{it}$ + β_3 *property repossessions*$_{it}$ + β_4 ln(*average wholesale bill*$_{it}$) + ε_{it}

B2. ln(*bad debt related operating costs*$_{it}$) = β_0 + β_1 ln(*total customers*$_{it}$) + β_2 *IMD income*$_{it}$ + β_3 ln(*average wholesale bill*$_{it}$) + ε_{it}

B3. ln(non-*bad debt related operating costs*$_{it}$) = β_0 + β_1 ln(*total customers*$_{it}$) + β_2 ln(*single service customers*$_{it}$) + β_3 *metered households*$_{it}$ + β_4 ln(peak traffic speed$_{it}$) + ε_{it}

B4. ln(*total retail operating costs*$_{it}$) = β_0 + β_1 ln(*total customers*$_{it}$) + β_2 ln(*single service customers*$_{it}$) + β_3 *metered properties*$_{it}$ + β_4 *IMD income*$_{it}$ + β_5 *property repossessions*$_{it}$ + β_6 ln(*average wholesale bill*$_{it}$) + ε_{it}

B5. ln(*total retail operating costs*$_{it}$) = β_0 + β_1 ln(*total customers*$_{it}$) + β_2 ln(*single service customers*$_{it}$) + β_3 *property repossessions*$_{it}$ + β_4 ln(*average wholesale bill*$_{it}$) + u_i + v_{it}

B6. ln(*bad debt related operating costs*$_{it}$) = β_0 + β_1 ln(*total customers*$_{it}$) + β_2 *IMD income*$_{it}$ + β_3 *property repossessions*$_{it}$ + β_4 ln(*average wholesale bill*$_{it}$) + u_i + v_{it}

B7. ln(*non-bad debt related operating costs*$_{it}$) = β_0 + β_1 ln(*total customers*$_{it}$) + β_2 ln(*single service customers*$_{it}$) + β_3 *metered households*$_{it}$ + β_4 ln(peak traffic speed$_{it}$) + β_5 time trend$_t$ + u_i + v_{it}

B8. B8. ln(*total retail operating costs*$_{it}$) = β_0 + β_1 ln(*total customers*$_{it}$) + β_2 ln(*single service customers*$_{it}$) + β_3 *metered households*$_{it}$ + β_4 *property repossessions*$_{it}$ + β_5 ln(*average wholesale bill*$_{it}$) + u_i + v_{it}

Description of dependent variable (eg what's excluded, gross v net etc)

Total retail operating costs: The totality of household operating retail costs, including opex and capital costs: customer services; debt management; doubtful debts; meter reading; services to developers; other operating expenditure; local authority rates; exceptional items; third party services; depreciation and amortisation.

Bad debt related retail operating costs: A subset of total retail operating costs, namely: debt management and doubtful debt.

Non-bad debt related retail operating costs: The subset of total retail operating costs not included in bad debt related retail operating costs – that is, all household retail operating costs other than debt management and doubtful debt.

Brief comment on models

A full description of the work undertaken to arrive at these models is set out in a report by Economic Insight: *'Household retail cost assessment for PR19: final report for Bristol and Wessex Water.'*

The models were developed using an objective general to specific methodology, which was subject to academic peer review. This generated a suite of 16 econometric models:

- Generalised models used a wide set of variables derived from a 'first principles' consideration of the drivers of retail costs.
- Specific models were estimated taking a 'liberal' approach to statistical significance (i.e. including variables that were significant at levels approaching 10%).
- 'Alternative' models were estimated for total retail operating costs, which retained variables that were not significant, but were correctly signed.
- Two approaches were used in the inclusion of scale (customer numbers) and scope (dual versus single service): Models A1 to A8 include separate variables for the number of dual and single service customers. Models B1 to B8 include a variable for total customer numbers, alongside the number of single service customers (where this remains after general to specific modelling).
- For each of the above, pooled OLS and random effects GLS models were estimated.

Overall, we consider the models across the suite to be valid. We regard pooled OLS and random effects GLS models as complementary: random effects models can distinguish between noise and other components of the error term, but unlike pooled OLS, produce time-invariant inefficiency estimates. Further, we think that both approaches to the incorporation of scale and scope are valid, and each has advantages and disadvantages. Using separate dual and single service variables provides a very flexible specification, and the resulting models incorporate a wider range of potentially relevant variables. On the other hand, the coefficients are difficult to interpret, as some companies have no dual service customers. The alternative approach is less flexible, but provides more intuitive coefficient estimates.

Note: p-values shown below are based on heteroscedasticity-robust standard errors for pooled OLS models (A1 to A4 and B1 to B4) and unadjusted standard errors for random effects GLS models (A5 to A8 and B5 to B8) – because this method already takes account of the correlation of errors within firms. This is consistent with the approach at PR14 and the approach we used in general to specific modelling.

	Model A1	Model A2	Model A3	Model A4	Model A5	Model A6	Model A7	Model A8
Dependent variable	ln(total retail operating costs)	ln(bad debt related operating costs)	ln(non-bad debt related operation costs)	ln(total retail operating costs)	ln(total retail operating costs)	ln(bad debt related operating costs)	ln(non-bad debt related operation costs)	ln(total retail operating costs)
ln(single service customers)	0.536*** (0.000)	0.535*** (0.000)	0.498*** (0.000)	0.563*** (0.0000)	0.349*** (0.001)	0.532*** (0.000)	0.268** (0.025)	0.318*** (0.003)
ln(dual service customers)	0.122*** (0.000)	0.121*** (0.000)	0.263*** (0.000)	0.159*** (0.0000)	0.226*** (0.000)	0.184*** (0.003)	0.250*** (0.000)	0.246*** (0.000)
Metered customers (%)			0.0143*** (0.0002)	0.00723* (0.062)			0.00214 (0.610)	0.00198 (0.500)
Metered household density (per km mains)			-0.0155*** (0.001)	-0.00662** (0.041)				
Flats (%)	0.0571*** (0.000)			0.0604*** (0.001)				0.0526 (0.144)
ln(peak traffic speed)			-1.830*** (0.0000)	-0.364 (0.290)			-1.217** (0.047)	
IMD income (%)	0.164*** (0.000)	0.189*** (0.000)		0.155*** (0.000)	0.0657 (0.167)	0.136*** (0.008)		0.105* (0.056)
Property repossessions (%)					0.107*** (0.000)			0.119*** (0.002)
ln(average wholesale bill)	1.206*** (0.000)	1.744*** (0.000)		0.999*** (0.000)	0.341 (0.000)	1.235*** (0.002)		0.301 (0.213)
Internal population total flow (%)		0.0909*** (0.001)						
Time trend							-0.0372** (0.014)	
Constant	-10.02*** (0.000)	-14.37*** (0.000)	4.539*** (0.000)	-8.063*** (0.000)	-2.741 (0.103)	-10.25*** (0.000)	4.104* (0.067)	-3.836** (0.039)
R2 adj (‘R2 overall’ in the case of random effects)	0.9284	0.9333	0.8743	0.9283	0.8957	0.9260	0.8539	0.9060

Reset test	0.0164	0.0004	0.0025	0.0061	0.0000	0.0030	0.0010	0.0000
VIF (max)	6.98	6.78	2.83	13.49	6.78	6.78	1.40	8.12
Method (eg OLS or RE)	Pooled OLS	Pooled OLS	Pooled OLS	Pooled OLS	Random effects GLS	Random effects GLS	Random effects GLS	Random effects GLS
N (sample size)	89	89	89	89	89	89	89	89

	Model B1	Model B2	Model B3	Model B4	Model B5	Model B6	Model B7	Model B8
Dependent variable	ln(total retail operating costs)	ln(bad debt related operating costs)	ln(non-bad debt related operation costs)	ln(total retail operating costs)	ln(total retail operating costs)	ln(bad debt related operating costs)	ln(non-bad debt related operation costs)	ln(total retail operating costs)
ln(total customers)	0.877*** (0.000)	0.979*** (0.000)	1.061*** (0.000)	0.966*** (0.000)	1.043*** (0.000)	0.933*** (0.000)	1.069*** (0.000)	1.065*** (0.000)
ln(single service customers)			-0.120*** (0.000)	-0.0690* (0.087)	-0.134** (0.041)		-0.138** (0.021)	-0.150** (0.030)
ln(dual service customers)								
Metered customers (%)			0.00452*** (0.004)	0.00473*** (0.005)			0.00461 (0.114)	0.00201 (0.400)
ln(peak traffic speed)			-0.257* (0.062)				-0.327 (0.286)	
IMD income (%)	0.0273*** (0.001)	0.0668*** (0.000)		0.0274*** (0.003)		0.0553* (0.071)		
Property repossessions (%)	0.121*** (0.000)			0.147*** (0.000)	0.113*** (0.000)	0.147** (0.015)		0.130*** (0.000)
ln(average wholesale bill)	0.659*** (0.000)	1.091*** (0.000)		0.480*** (0.000)	0.400*** (0.004)	1.165*** (0.000)		0.351** (0.019)
Internal population total flow (%)								
Time trend							-0.0349*** (0.002)	
Constant	-6.974*** (0.000)	-11.31*** (0.000)	-3.200*** (0.000)	-6.502*** (0.0000)	-5.519*** (0.000)	-11.57*** (0.000)	-2.820** (0.011)	-5.446*** (0.000)
R2 adj ('R2 overall' in the case of random effects)	0.9821	0.9616	0.9676	0.9835	0.9815	0.9639	0.9709	0.9824
Reset test	0.2036	0.0308	0.0273	0.4076	0.0071	0.0174	0.0076	0.0169

VIF	2.62	2.07	1.44	9.81	5.79	2.62	1.44	7.84
Method (eg OLS or RE)	Pooled OLS	Pooled OLS	Pooled OLS	Pooled OLS	Random effects GLS	Random effects GLS	Random effects GLS	Random effects GLS
N (sample size)	89	89	89	89	89	89	89	89

Chart

Percentage variance between modelled and actual costs. Calculated as:

% variance = (ΣF – ΣA) / ΣF (F=Forecast; A=Actual)

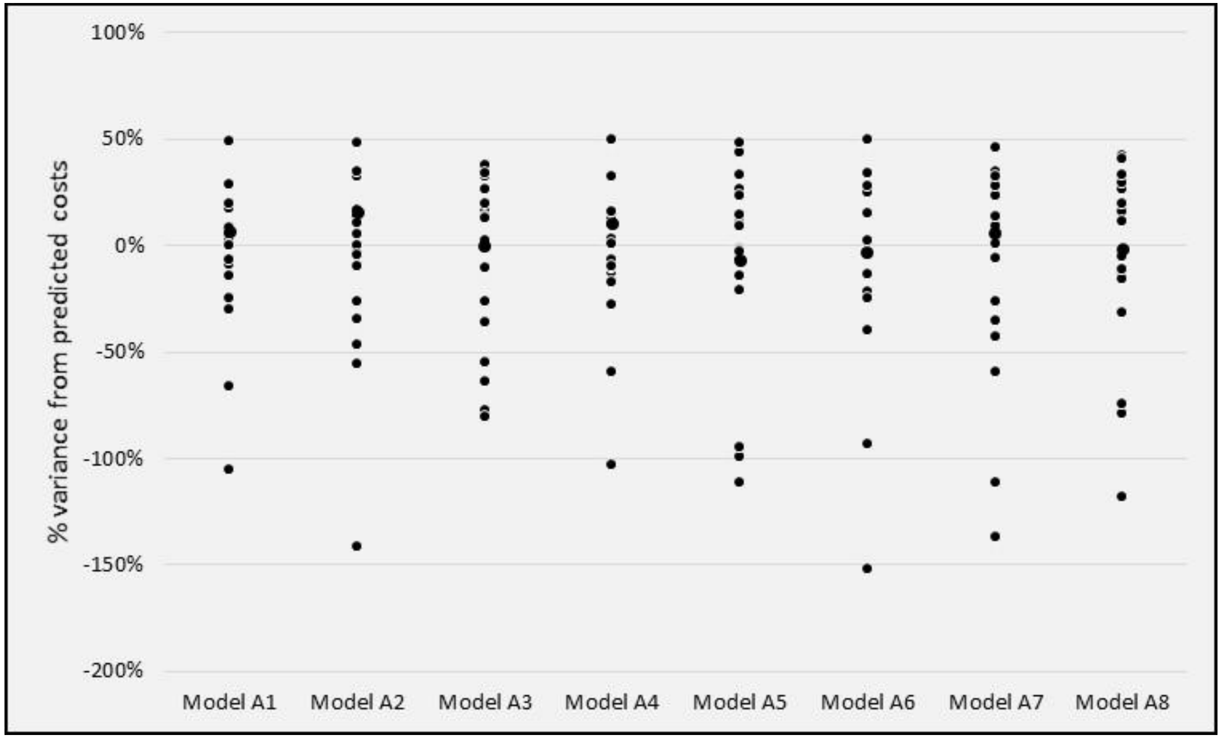
Where F and A are in pounds (not logs!) and the summation is over the sample period.

See Anglian Water's report on cost assessment, page 85

Model set A

Company	Model A1	Model A2	Model A3	Model A4	Model A5	Model A6	Model A7	Model A8
ANH	-14%	-9%	26%	-6%	11%	2%	9%	-13%
NES	49%	48%	37%	50%	44%	50%	24%	42%
NWT	-105%	-142%	-77%	-103%	-99%	-152%	-59%	-79%
SRN	3%	-26%	-55%	-13%	-21%	-22%	-6%	-4%
SVT	-25%	14%	-11%	-17%	-2%	25%	-35%	-16%
SWB	0%	33%	16%	3%	26%	25%	30%	26%

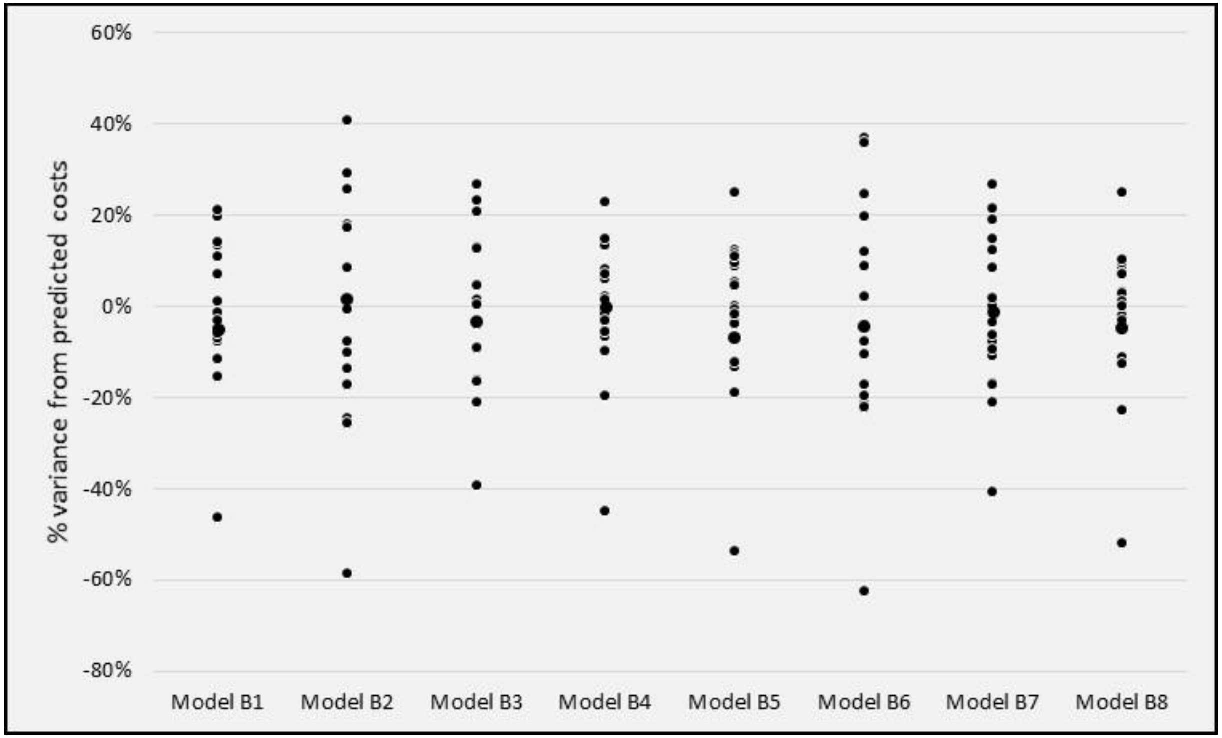
TMS	8%	11%	20%	10%	-7%	-22%	14%	16%
WSH	8%	-1%	-36%	1%	-2%	-3%	-26%	-5%
WSX	19%	0%	37%	12%	33%	15%	46%	19%
YKY	5%	17%	2%	10%	9%	25%	1%	12%
AFW	-30%	-56%	-80%	-28%	-111%	-93%	-137%	-74%
BRL	17%	-5%	32%	16%	-3%	-25%	24%	-11%
DVW	20%	11%	14%	16%	49%	34%	35%	41%
PRT	-9%	-46%	13%	-10%	24%	-13%	28%	29%
SES	29%	35%	34%	32%	15%	28%	32%	33%
SEW	-66%	-34%	-64%	-59%	-94%	-40%	-111%	-118%
SSC	-7%	5%	-26%	-10%	-14%	2%	-43%	-32%
SWT	7%	16%	0%	11%	-7%	-3%	6%	-2%



Model set B

Company	Model B1	Model B2	Model B3	Model B4	Model B5	Model B6	Model B7	Model B8
ANH	7%	-8%	21%	13%	5%	-10%	19%	9%
NES	20%	18%	1%	15%	9%	20%	0%	7%
NWT	-15%	-24%	5%	-10%	-2%	-22%	8%	-2%
SRN	-46%	-58%	-39%	-45%	-54%	-62%	-41%	-52%
SVT	1%	29%	-16%	-1%	4%	25%	-17%	3%
SWB	-5%	9%	0%	2%	0%	9%	2%	3%
TMS	-2%	-26%	-4%	-2%	-4%	-8%	-3%	-3%

WSH	-12%	-17%	-21%	-20%	-19%	-20%	-21%	-23%
WSX	13%	-10%	23%	6%	12%	-17%	22%	10%
YKY	21%	41%	13%	23%	25%	37%	15%	25%
AFW	-2%	-10%	-10%	1%	-1%	2%	-11%	1%
BRL	14%	-14%	27%	8%	12%	-21%	27%	9%
DVW	-8%	-26%	-16%	-3%	-13%	-20%	-17%	-11%
PRT	-3%	-17%	13%	-7%	10%	-22%	12%	7%
SES	11%	26%	-10%	7%	11%	36%	-8%	10%
SEW	-7%	17%	-4%	-3%	-2%	12%	-6%	0%
SSC	-6%	-1%	-9%	-5%	-12%	-4%	-9%	-12%
SWT	-5%	1%	-3%	0%	-7%	-5%	-1%	-5%



Comments:

- Please indicate the units of the explanatory variable, and whether it was expressed in logs.
- Use asterisks to denote significance level: *** (1%), ** (5%) and * (10%)
- P values should be based on cluster robust standard errors
- In the case of random effects please report Stata's output "R2 overall"

Template for submission of econometric models for consultation
Sewage treatment models 1

Econometric model formula:

1. $\ln(Expenditure) = \propto_0 + \propto_1 \ln\big(Total\ load(BOD)\big) + \propto_2 \ln(Average\ size\ of\ works) + \propto_3\ (proportion\ of load\ undergoing\ tertiary\ treatment) + time\ dummy\ variables$

2. $\ln(Smoothed\ expenditure) = \propto_0 + \propto_1 \ln\big(Total\ load(BOD)\big) + \propto_2 \ln(Average\ size\ of\ works) + \propto_3\ (proportion\ of load\ undergoing\ tertiary\ treatment) + time\ dummy\ variables$

3. $\ln\left(\frac{Expenditure}{sludge\ produced}\right) = \propto_0 + \propto_1 \ln(Average\ size\ of\ works) + \propto_2\ (proportion\ of load\ undergoing\ tertiary\ treatment) + + time\ dummy\ variables$

4. $\ln\left(\frac{Smoothed\ expenditure}{Sludge\ produced}\right) = \propto_0 + \propto_1 \ln(Average\ size\ of\ works) + \propto_2\ (proportion\ of load\ undergoing\ tertiary\ treatment) + time\ dummy\ variables$

Description of dependent variable

Model 1 Sewage Treatment botex := Opex + capital Maintenance – third party costs – local authority rates – EA charges

Model 2 Sewage Treatment botex := Opex + IRE + average MNI over period – third party costs – local authority rates – EA charges

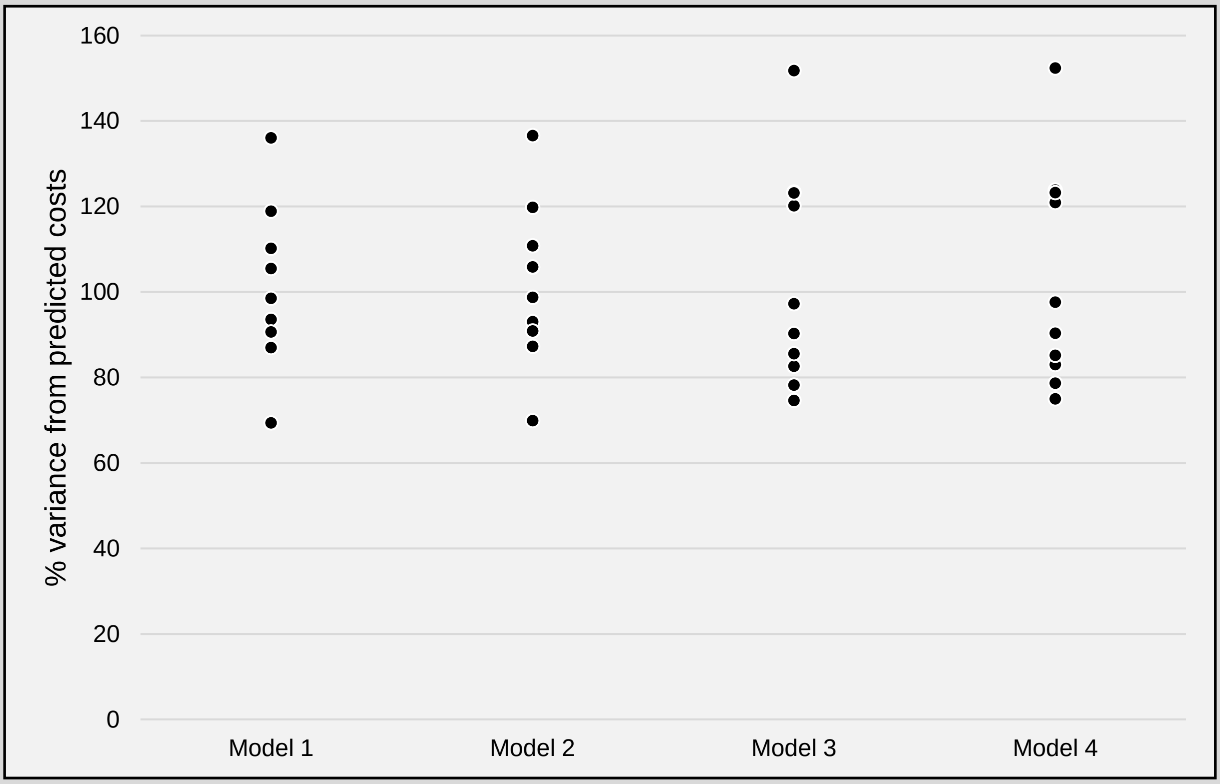
Brief comment on models

These are our Endogenous STW models. Please see attached not for more information.

Ofwat measure of highly dense areas := the proportion of the companies area of service with over 6000 pop. Density as defined in constructed dataset on the Ofwat SharePoint.

	Model 1	**Model 2**	**Model 3**	**Model 4**
Dependent variable	Aggregate Sewage Treatment Botex	Smoothed Sewage Treattment Botex	Unit Sewage Treatment Botex	Smoothed Unit Sewage Treatment Botex
Total load (BOD)	0.7140*** (0.000)	0.7100*** (0.000)		
Average size of works (total load / total works)	0.0424 (0.177)	0.0448 (0.137)	-0.0086 (0.809)	-0.0070 (0.841)
Proportion of load undergoing tertiary treatment	0.0605 (0.883)	0.0655 (0.878)	0.3469 (0.540)	0.3559 (0.535)
Constant	-4.4611** (0.013)	-4.4516** (0.014)	-7.9925*** (0.000)	-8.032*** (0.000)
R2 adjusted	0.8605	0.8782	0.0764	0.0679
VIF (max)	1.71	1.71	1.69	1.69
Reset test	0.000	0.000	0.5808	0.1108
Estimation method (eg OLS or RE)	OLS	OLS	OLS	OLS
N (sample size)	60	60	60	60

Chart


160
140
120
100
80
60
40
20
0
% variance from predicted costs
Model 1
Model 2
Model 3
Model 4

Template for submission of econometric models for consultation
Sewage treatment models 2

Econometric model formula:

1. $\ln(Expenditure) = \propto_0 + \propto_1 \ln\big(Total\ load(BOD)\big) + \propto_2\ (Ofwat\ measure\ of\ highly\ dense\ areas) + \propto_3\ (proportion\ of load\ undergoing\ tertiary\ treatment) + time\ dummy\ variables$

2. $\ln(Smoothed\ expenditure) = \propto_0 + \propto_1 \ln\big(Total\ load(BOD)\big) + \propto_2\ (Ofwat\ measure\ of\ highly\ dense\ areas) + \propto_3\ (proportion\ of load\ undergoing\ tertiary\ treatment) + time\ dummy\ variables$

3. $\ln\left(\frac{Expenditure}{sludge\ produced}\right) = \propto_0 + \propto_1\ (Ofwat\ measure\ of\ highly\ dense\ areas) + \propto_2\ (proportion\ of load\ undergoing\ tertiary\ treatment) + + time\ dummy\ variables$

4. $\ln\left(\frac{Smoothed\ expenditure}{Sludge\ produced}\right) = \propto_0 + \propto_1\ (Ofwat\ measure\ of\ highly\ dense\ areas) + \propto_2\ (proportion\ of load\ undergoing\ tertiary\ treatment) + time\ dummy\ variables$

Description of dependent variable

Model 1 Sewage Treatment botex := Opex + capital Maintenance – third party costs – local authority rates – EA charges

Model 2 Sewage Treatment botex := Opex + IRE + average MNI over period – third party costs – local authority rates – EA charges

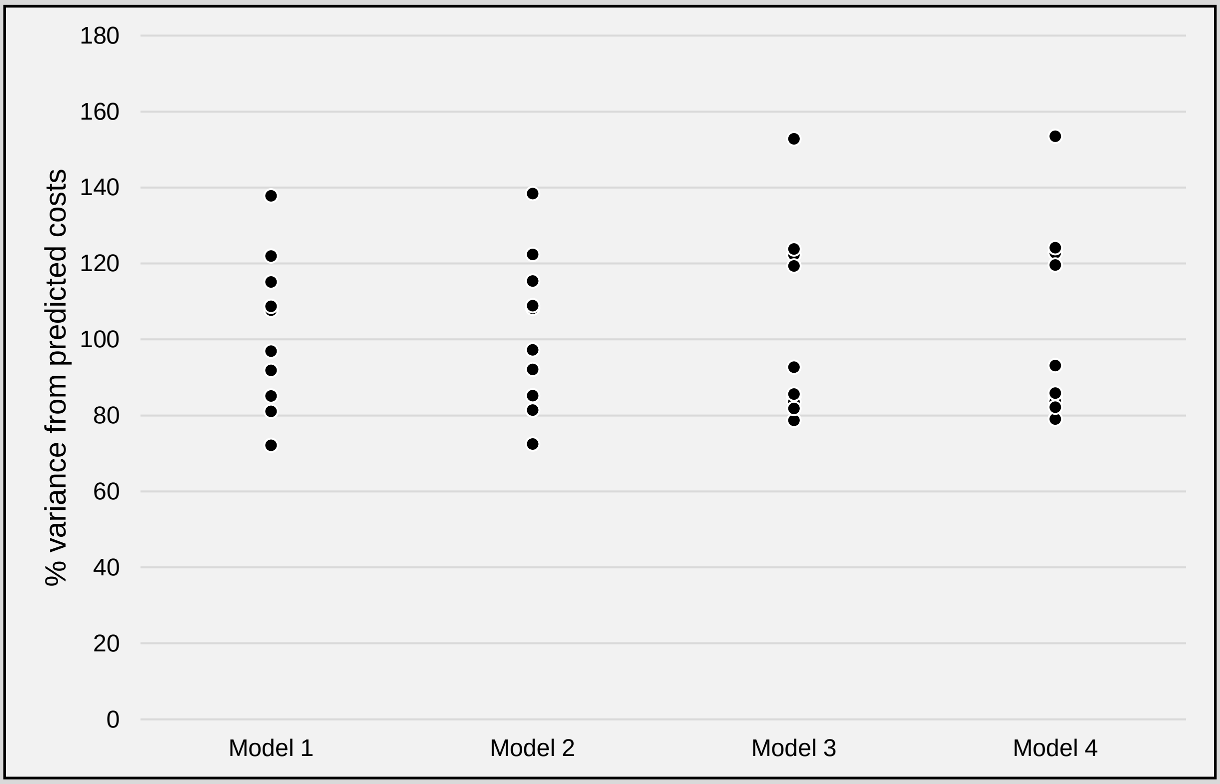
Brief comment on models

These are our Exogenous STW models. Please see attached not for more information.

Ofwat measure of highly dense areas := the proportion of the companies area of service with over 6000 pop. Density as defined in constructed dataset on the Ofwat SharePoint.

	Model 1	**Model 2**	**Model 3**	**Model 4**
Dependent variable	Aggregate Sewage Treatment Botex	Smoothed Sewage Treattment Botex	Unit Sewage Treatment Botex	Smoothed Unit Sewage Treatment Botex
Total load (BOD)	0.7570*** (0.000)	0.7581*** (0.000)		
Ofwat measure of highly dense areas	0.1468 (0.680)	0.1418 (0.691)	-0.2645 (0.407)	-0.2678 (0.404)
Proportion of load undergoing tertiary treatment	0.0336 (0.936)	0.0423 (0.923)	0.3974 (0.465)	0.4046 (0.463)
Constant	-4.8088** (0.022)	-4.8523** (0.023)	-7.9930*** (0.000)	-8.022*** (0.000)
R2 adjusted	0.8477	0.8632	0.1190	0.1179
VIF (max)	1.70	1.70	1.68	1.68
Reset test	0.000	0.000	0.0398	0.0036
Estimation method (eg OLS or RE)	OLS	OLS	OLS	OLS
N (sample size)	60	60	60	60

Chart


180
160
140
120
100
80
60
40
20
0
% variance from predicted costs
Model 1
Model 2
Model 3
Model 4

Template for submission of econometric models for consultation
Sewerage models

Econometric model formula:

1. $\ln(Expenditure) = \propto_0 + \propto_1 \ln(Connected\ Properties) + \propto_2 \ln(density) + \propto_3 \ln(density)^2 + time\ dummy\ variables$

2. $\ln(Smoothed\ expenditure) = \propto_0 + \propto_1 \ln(connected\ properties) + \propto_2 \ln(density) + \propto_3 \ln(density)^2 + time\ dummy\ variables$

3. $\ln\left(\frac{Expenditure}{properties}\right) = \propto_0 + \propto_1 \ln(density) + \propto_2 \ln(density)^2\ time\ dummy\ variables$

4. $\ln\left(\frac{Smoothed\ expenditure}{properties}\right) = \propto_0 + \propto_1 \ln(density) + \propto_2 \ln(density)^2 + time\ dummy\ variables$

Description of dependent variable

Model 2 & 4 – sewerage botex := Opex + capital Maintenance – third party costs – local authority rates – abstraction charges

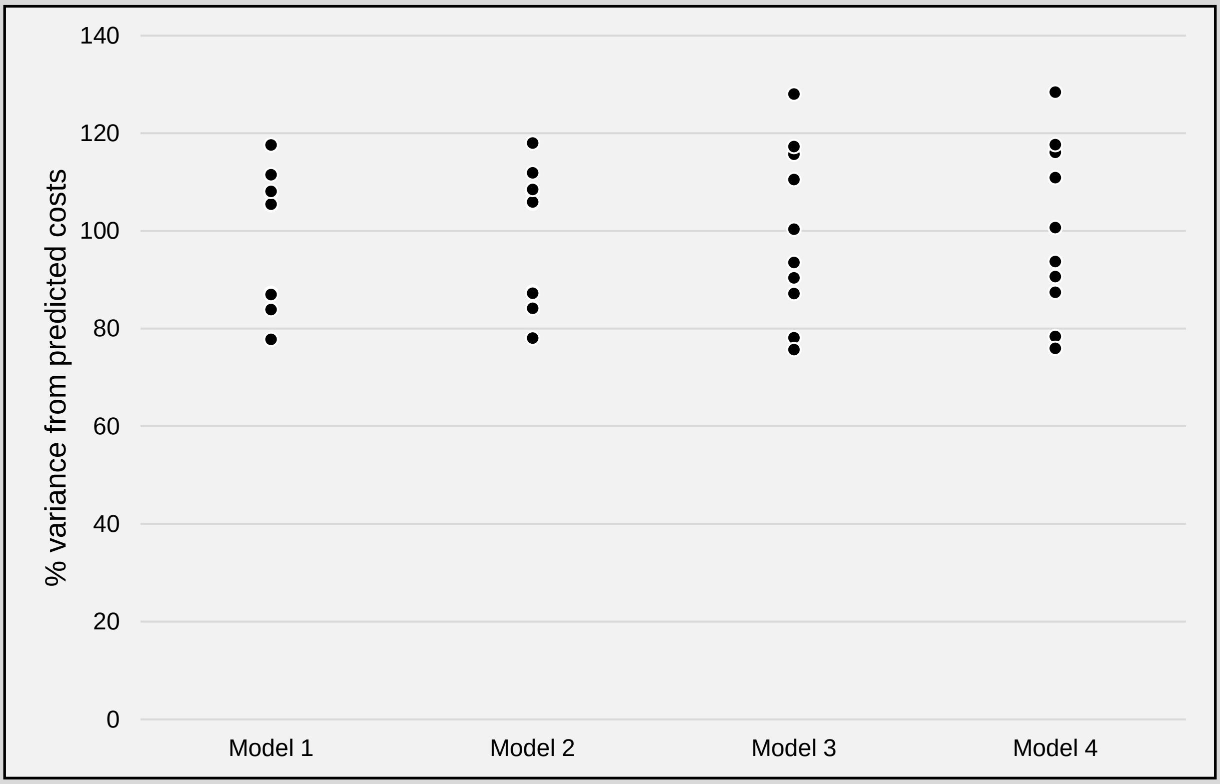
Model 1 & 3 - sewerage botex := Opex + IRE + average MNI over period – third party costs – local authority rates – abstraction charges

Brief comment on models

Please see our attached note for full details of the modelling.

	Model 1	Model 2	Model 3	Model 4
Dependent variable	Smoothed Aggregate SC Botex	Aggregate SC Botex	Smoothed Unit SC Botex	Unit SC Botex
Connected Properties	0.6838*** (0.000)	0.6852*** (0.000)		
Sewage Catchment area per 1k properties	-0.2122 (0.142)	-0.2103 (0.144)	0.0350 (0.777)	0.0359 (0.771)
Sewage Catchment area per 1k properties ^2	-0.1417 (0.697)	-0.1430 (0.695)	-0.6557** (0.075)	-0.6547** (0.075)
Constant	-0.6307 (0.430)	-0.6387 (0.421)	-2.9306*** (0.000)	-2.9286*** (0.000)
R2 adjusted	0.9002	0.8946	0.3369	0.3275
VIF (max)	3.48	3.48	1.67	1.67
Reset test	0.0252	0.0242	0.0002	0.0004
Estimation method (eg OLS or RE)	OLS	OLS	OLS	OLS
N (sample size)	60	60	60	60

Chart


% variance from predicted costs
140
120
100
80
60
40
20
0
Model 1
Model 2
Model 3
Model 4

Template for submission of econometric models for consultation
Water distribution models

Econometric model formula:

1. $\ln(Expenditure) = \propto_0 + \propto_1 \ln(Connected\ Properties) + \propto_2 \ln(SR\ per\ 100k\ properties) + \propto_3 \ln(SR\ per\ 100k\ properties)^2 + \propto_4 \ln(Average\ pumping\ head) + time\ dummy\ variables$

2. $\ln(Smoothed\ expenditure) = \propto_0 + \propto_1 \ln(Connected\ Properties) + \propto_2 \ln(SR\ per\ 100k\ properties) + \propto_3 \ln(SR\ per\ 100k\ properties)^2 + \propto_4 \ln(Average\ pumping\ head) + time\ dummy\ variables$

3. $\ln\left(\frac{Expenditure}{properties}\right) = \propto_0 + \propto_1 \ln(SR\ per\ 100k\ properties) + \propto_2 \ln(SR\ per\ 100k\ properties)^2 + \propto_3 \ln(Average\ pumping\ head) + time\ dummy\ variables$

4. $\ln\left(\frac{Smoothed\ expenditure}{properties}\right) = \propto_0 + \propto_1 \ln(SR\ per\ 100k\ properties) + \propto_2 \ln(SR\ per\ 100k\ properties)^2 + \propto_3 \ln(Average\ pumping\ head) + time\ dummy\ variables$

Description of dependent variable

Model 2 & 4 – Water treatment and resources botex := Opex + capital Maintenance – third party costs – local authority rates – abstraction charges

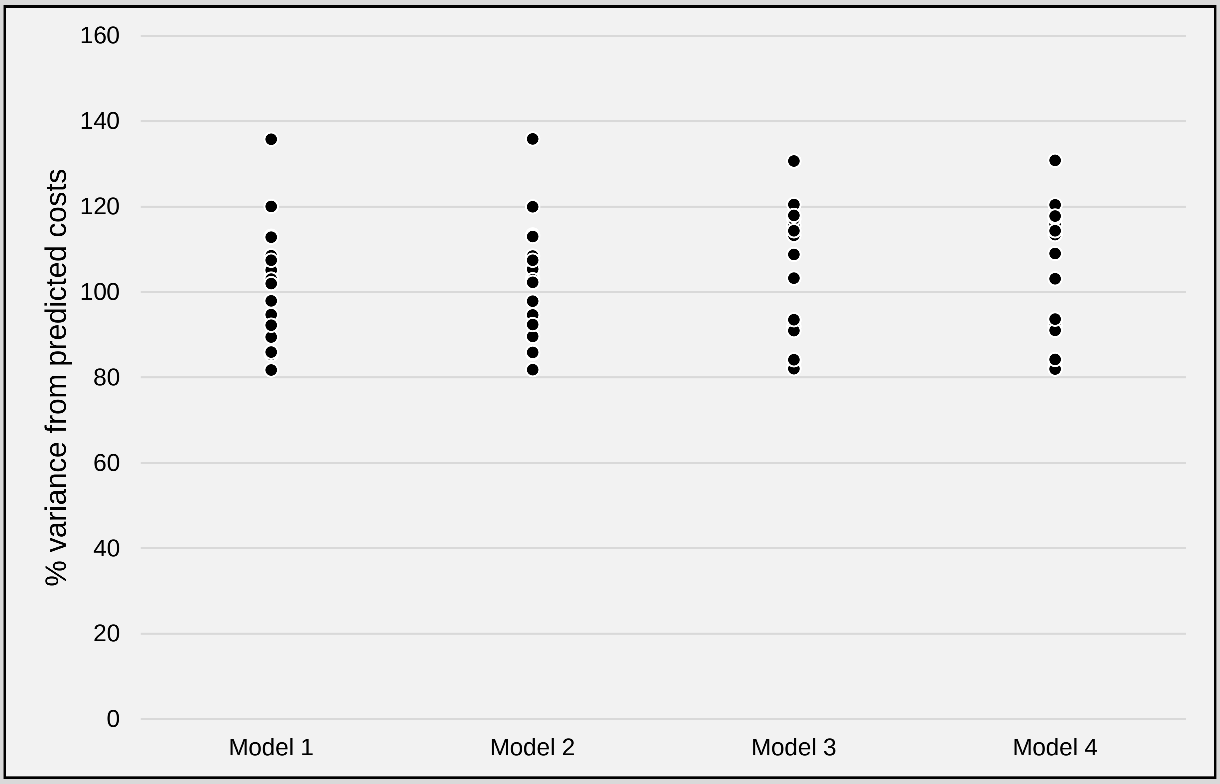
Model 1 & 3 - Water treatment and resources botex := Opex + IRE + average MNI over period – third party costs – local authority rates – abstraction charges

Brief comment on models

These are our water distribution models, outlined further in our attached note

	Model 1	Model 2	Model 3	Model 4
Dependent variable	Smoothed Aggregate WD Botex	Aggregate WD Botex	Smoothed Unit WD Botex	Unit WD Botex
Connected Properties	1.0875*** (0.000)	1.0866*** (0.000)		
Service Reservoirs / 100k properties	-2.9959*** (0.000)	-3.0153*** (0.000)	-2.9093*** (0.000)	-2.9296*** (0.000)
Service Reservoirs / 100k properties ^ 2	0.5535*** (0.000)	0.5569*** (0.000)	0.5291*** (0.000)	0.5327*** (0.000)
Average pumping head	0.1493* (0.053)	0.1497* (0.052)	0.1284 (0.149)	0.190 (0.146)
Constant	-0.0664 (0.928)	-0.090 (0.979)	0.5830 (0.519)	0.6234 (0.487)
R2 adjusted	0.9739	0.9698	0.4573	0.4389
VIF (max)	111.25 Note – driven by including quadratic terms	111.25 Note – driven by including quadratic terms	110.25 Note – driven by including quadratic terms	110.25 Note – driven by including quadratic terms
Reset test	0.3340	0.4848	0.0062	0.0366
Estimation method (eg OLS or RE)	OLS	OLS	OLS	OLS
N (sample size)	102	102	102	102

Chart


160
140
120
100
80
60
40
20
0
% variance from predicted costs
Model 1
Model 2
Model 3
Model 4

Template for submission of econometric models for consultation
Water treatment models 1

Econometric model formula:

1. $\ln(Expenditure) = \propto_0 + \propto_1 \ln\big(distirbution\ inputs\ (DI)\big) + \propto_2\ (Ofwat\ measure\ of\ highly\ dense\ areas) + \propto_3\ (Proprotio\ of\ DI\ from\ groundwater) + \propto_4 \ln(Average\ pumping\ head) + time\ dummy\ variables$

2. $\ln(Smoothed\ expenditure) = \propto_0 + \propto_1 \ln\big(distirbution\ inputs\ (DI)\big) + \propto_2\ (Ofwat\ measure\ of\ highly\ dense\ areas) + \propto_3\ (Proprotio\ of\ DI\ from\ groundwater) + \propto_4 \ln(Average\ pumping\ head) + time\ dummy\ variables$

3. $\ln\frac{(Expenditure)}{DI} = \propto_0 + \propto_1\ (Ofwat\ measure\ of\ highly\ dense\ areas) + \propto_2\ (Proprotio\ of\ DI\ from\ groundwater) + \propto_3 \ln(Average\ pumping\ head) + time\ dummy\ variables$

4. $\ln\left(\frac{Smoothed\ expenditure}{DI}\right) = \propto_0 + \propto_1\ (Ofwat\ measure\ of\ highly\ dense\ areas) + \propto_2\ (Proprotio\ of\ DI\ from\ groundwater) + \propto_3 \ln(Average\ pumping\ head) + time\ dummy\ variables$

Description of dependent variable

Model 1 & 3 – Water treatment and resources botex := Opex + capital Maintenance – third party costs – local authority rates – abstraction charges

Model 2 & 4 - Water treatment and resources botex := Opex + IRE + average MNI over period – third party costs – local authority rates – abstraction charges

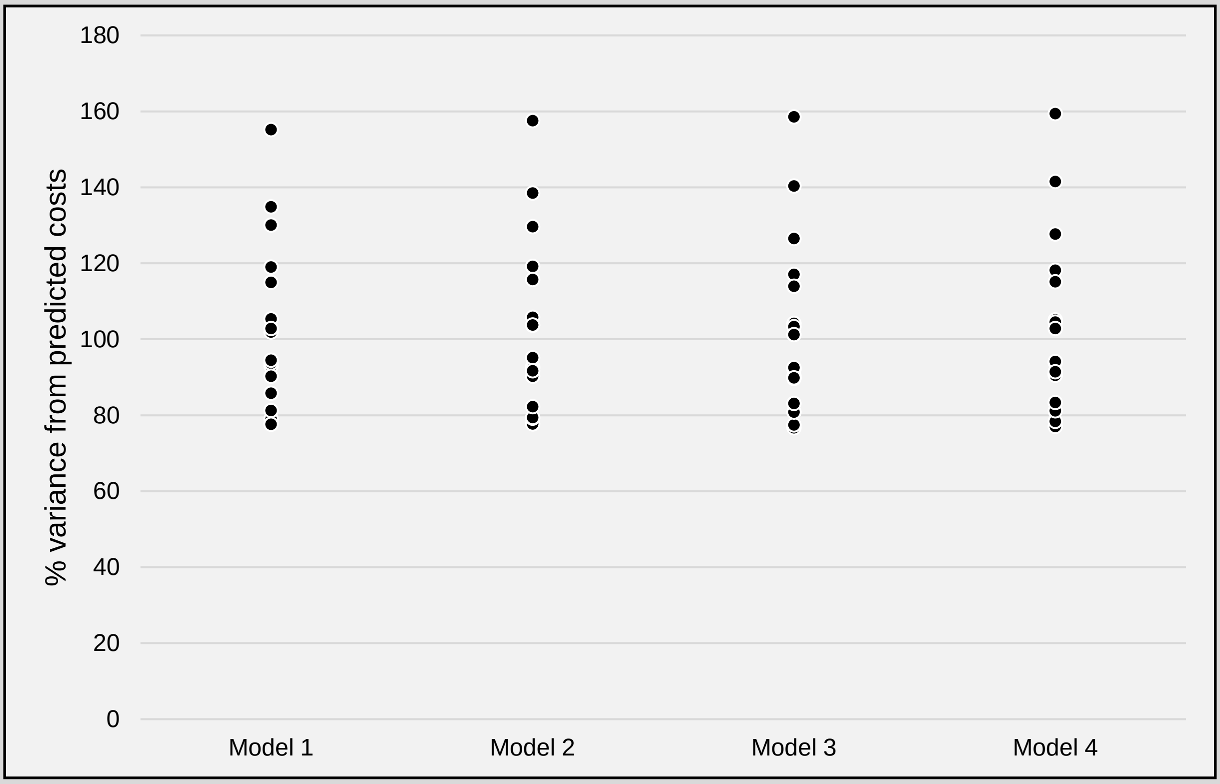
Brief comment on models

These are the exogenous variations of our water treatment & resource models.

Ofwat measure of highly dense areas := the proportion of the companies area of service with over 6000 pop. Density as defined in constructed dataset on the Ofwat SharePoint.

	Model 1	**Model 2**	**Model 3**	**Model 4**
Dependent variable	Aggregate WT&R Botex	Smoothed Aggregate WT&R Botex	Unit WT&R Botex	Smoothed Unit WT&R Botex
Distribution Input	0.9810*** (0.000)	0.9653*** (0.000)		
Ofwat measure of highly dense areas	-0.5326** (0.019)	-0.5453** (0.017)	-0.5435** (0.014)	-0.5653** (0.013)
Proportion of DI from groundwater sources	-0.2742 (0.131)	-0.2955* (0.099)	-0.2518 (0.164)	-0.2547 (0.165)
Average pumping head	0.2692*** (0.006)	0.2611*** (0.008)	0.2794*** (0.006)	0.2798*** (0.008)
Constant	-3.089*** (0.000)	-2.948*** (0.000)	-3.2534*** (0.000)	-3.2489*** (0.000)
R2 adjusted	0.9376	0.9538	0.5066	0.5905
VIF (max)	1.67	1.67	1.67	1.67
Reset test	0.0008	0.0006	0.000	0.000
Estimation method (eg OLS or RE)	OLS	OLS	OLS	OLS
N (sample size)	102	102	102	102

Chart


180
160
140
120
100
80
60
40
20
0
% variance from predicted costs
Model 1
Model 2
Model 3
Model 4

Template for submission of econometric models for consultation
Water treatment models 2

Econometric model formula:

1. $\ln(Expenditure) = \propto_0 + \propto_1 \ln\big(distirbution\ input\ (DI)\big) + \propto_2 \ln(Average\ size\ of\ source) + \propto_3\ (Proprotion\ of\ DI\ treated\ at\ W4\ sites +) + \propto_4 \ln(Average\ pumping\ head) + time\ dummy\ variables$

2. $\ln(Smoothed\ expenditure) = \propto_0 + \propto_1 \ln\big(distirbution\ input\ (DI)\big) + \propto_2 \ln(Average\ size\ of\ source) + \propto_3\ (Proprotion\ of\ DI\ treated\ at\ W4\ sites +) + \propto_4 \ln(Average\ pumping\ head) + time\ dummy\ variables$

3. $\ln\left(\frac{Expenditure}{DI}\right) = \propto_0 + \propto_1 \ln(Average\ size\ of\ source) + \propto_2\ (Proprotion\ of\ DI\ treated\ at\ W4\ sites +) + \propto_3 \ln(Average\ pumping\ head) + time\ dummy\ variables$

4. $\ln\left(\frac{Smoothed\ expenditure}{DI}\right) = \propto_0 + \propto_1 \ln(Average\ size\ of\ source) + \propto_2\ (Proprotion\ of\ DI\ treated\ at\ W4\ sites +) + \propto_3 \ln(Average\ pumping\ head) + time\ dummy\ variables$

Description of dependent variable

Model 1 & 3 – Water treatment and resources botex := Opex + capital Maintenance – third party costs – local authority rates – abstraction charges

Model 2 & 4 - Water treatment and resources botex := Opex + IRE + average MNI over period – third party costs – local authority rates – abstraction charges

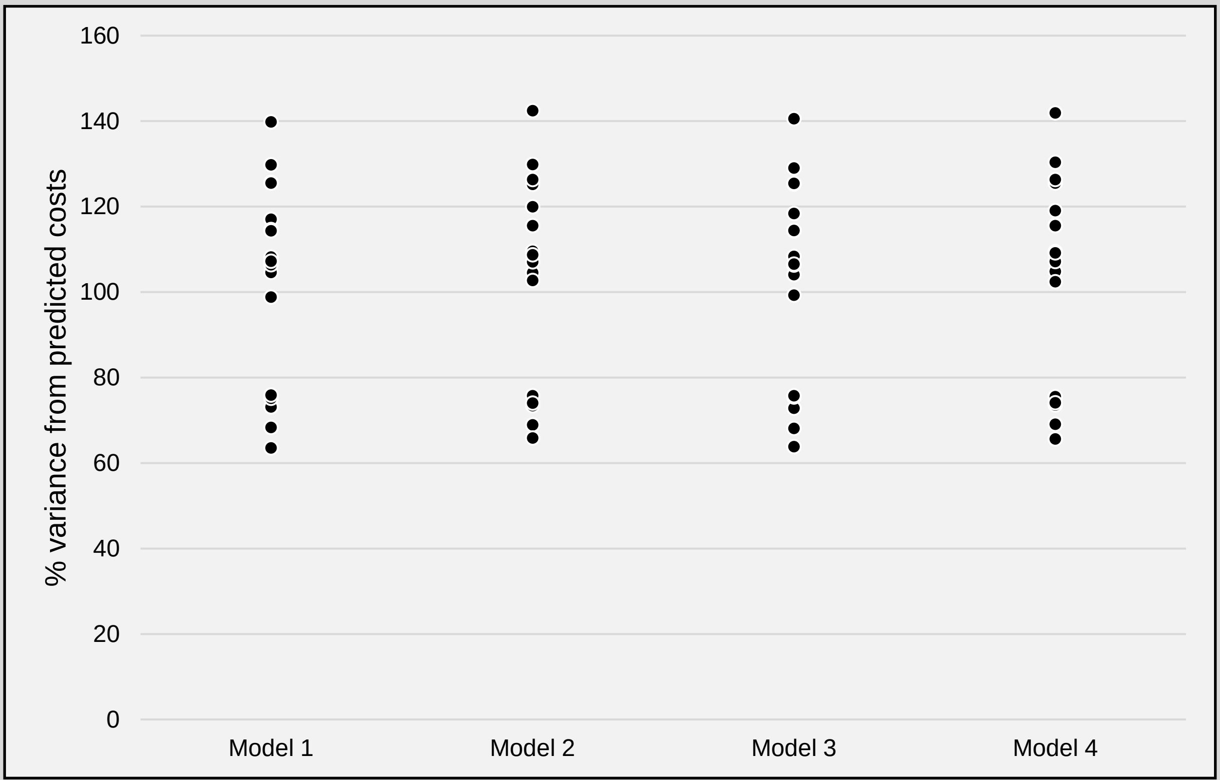
Brief comment on models

These are the endogenous variations of our water treatment & resource models.

Ofwat measure of highly dense areas := the proportion of the companies area of service with over 6000 pop. Density as defined in constructed dataset on the Ofwat SharePoint.

	Model 1	**Model 2**	**Model 3**	**Model 4**
Dependent variable	Aggregate WT&R Botex	Smoothed Aggregate WT&R Botex	Unit WT&R Botex	Smoothed Unit WT&R Botex
Distribution Input	1.0037*** (0.000)	0.9941*** (0.000)		
Average Source Size	-0.1171 (0.719)	-0.1669 (0.587)	-0.1092 (0.759)	-0.1793 (0.585)
Proportion of water treated W4+	0.0132 (0.322)	0.0158 (0.211)	0.0129 (0.370)	0.0162 (0.219)
Average pumping head	0.4124*** (0.003)	0.4108*** (0.004)	0.4104*** (0.001)	0.4139*** (0.001)
Constant	-4.8758*** (0.000)	-4.8995*** (0.000)	-4.8390*** (0.000)	-4.9571*** (0.000)
R2 adjusted	0.9208	0.9352	0.3763	0.4366
VIF (max)	12.73	12.73	10.26	10.26
Reset test	0.0214	0.0125	0.6859	0.1674
Estimation method (eg OLS or RE)	OLS	OLS	OLS	OLS
N (sample size)	102	102	102	102

Chart


160
140
120
100
80
60
40
20
0
% variance from predicted costs
Model 1
Model 2
Model 3
Model 4