Econometric model formula:

1. **Total Botex Model 1:** Ln(total botex per connected property aggregate) = a + b1*Ln(DI/connected property) + b2*Ln(length of mains/ connected property) + b3*Ln(length of raw mains and conveyors/DI) + b4*Share of water treated at level 5 and above + b5*Length of mains laid pre-1940/Total length of mains + b6*Length of renewed and relined mains/Total length of mains + b7*Ln(average pumping head aggregate) + b8*Year15 + b9*Year16 + b10*Year17

2. **Total Botex Model 2:** Ln(total botex per connected property aggregate) =a + b1*Ln(DI/ connected property) + b2*Ln(length of mains/ connected property) + b3* Surface water treated / Total water treated + b4*Length of mains laid pre-1940/Total length of mains + b5*Length of renewed and relined mains/Total length of mains + b6*Share of water from reservoirs + b7* Ln(average pumping head aggregate) + b8*Ln(number of sources / DI) + b9*Year15 + b10*Year16 + b11*Year17

3. **Total Botex Model 3:** Ln(total botex per connected property aggregate) = a + b1*Ln(DI/connected property) + b2*Ln(length of mains/ connected property) + b3* Share of water treated at level 5 and above + b4* Length of mains laid pre-1940/Total length of mains + b5* Length of renewed and relined mains/Total length of mains + b6*Share of water from reservoirs + b7* Ln(average pumping head aggregate) + b8*Year15 + b9*Year16 + b10*Year17

4. **Water Resources Botex Model 1:** Ln(water resources botex per connected property resources) = a + b1*Share of water from boreholes + b2*Share of water from reservoirs + b3*Ln(number of sources / DI) + b4*Year15 + b5*Year16 + b6*Year17

5. **Water Resources Total Botex Model 2:** Ln(water resources botex per connected property resources) = a + b1*Ln(length of raw mains and conveyors/DI) + b2*Ln(average pumping head resources) + b3* Share of water from boreholes + b4*Ln(number of sources / DI) + b5* Share of water from reservoirs + b6*Year15 + b7*Year16 + b8*Year17

6. **Water Resources Total Botex Model 3:** Ln(water resources botex per connected property resources) = a + b1*Ln(length of raw mains and conveyors/DI) + b2*Ln(average pumping head resources) + b3* Share of water from reservoirs + b4*Ln(number of sources / DI) + b5*Year15 + b6*Year16 + b7*Year17

7. **Network Plus Total Botex Model 1:** Ln(network plus botex per connected property network) = a + b1*Ln(DI/connected property) + b2*Ln(length of mains/ connected property) + b3* Share of water treated at level 5 and above + b4* Length of mains laid pre-1940/Total length of mains + b5* Length of renewed and relined mains/Total length of mains + b6*Year15 + b7*Year16 + b8*Year17

8. **Network Plus Total Botex Model 2:** Ln(network plus botex per connected property network) = a + b1* Ln(length of mains/ connected property) + b2* Ln(DI/connected property) + b3* Share of water treated at level 5 and above + b4* Share of water from reservoirs + b5* Length of mains laid pre-1940/Total length of mains + b6* Length of renewed and relined mains/Total length of mains + b7* Ln(average pumping head network) + b8*Year15 +

b9*Year16 + b10*Year17

9. **Network Plus Total Botex Model 3:** Ln(network plus botex per connected property network) = a + b1*Ln(DI/connected property) + b2*Ln(length of mains/ connected property) + b3* Surface water treated / Total water treated + b4* Length of mains laid pre-1940/Total length of mains + b5* Length of renewed and relined mains/Total length of mains + b6*Ln(number of sources / DI) + b7*Year15 + b8*Year16 + b9*Year17

Description of dependent variable

Botex per connected property by value chain element (Water Resources, Network plus, Aggregate Wholesale) has been used as the dependent variable in the unit cost models presented in this pro forma. For each respective value chain element, botex has been calculated as the sum of opex and capital maintenance.

In reference to company submissions for the six-year wholesale cost data (July 2017), opex is made up of power, income treated as negative expenditure, abstraction charges, bulk supply costs, renewals expensed in year (infrastructure), renewals expensed in year (non-infrastructure) and other operating expenditure excluding renewals costs. Local authority and cumulo rates and third party services are not included. Capital maintenance is made up of maintaining the long term capability of assets (infra) + maintaining the long term capability of assets (non-infra) costs. Capital maintenance costs have been smoothed on a three year rolling-average basis, therefore four years of data have been modelled (2014-2017). Botex costs have been calculated on a unit cost basis by dividing cost information by the sum of Total non-household connected properties at year end and Total household connected properties at year end also from the six-year wholesale cost data set.

Brief comment on models

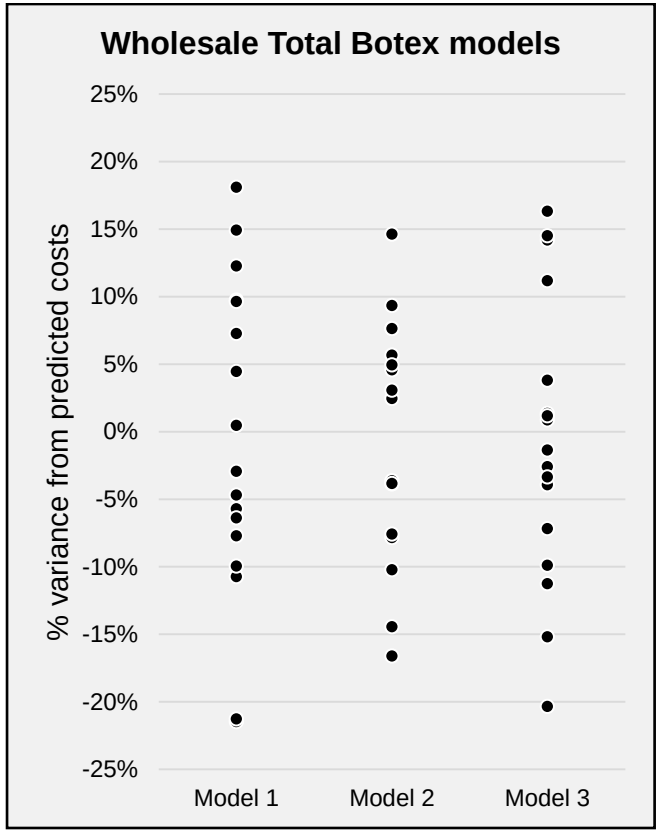
A full description of the work undertaken to arrive at these models is set out in a report by NERA: *'Comparative Benchmarking Assessment to Support Preparation of Bristol Water's AMP7 Business Plan'* (December 2017).

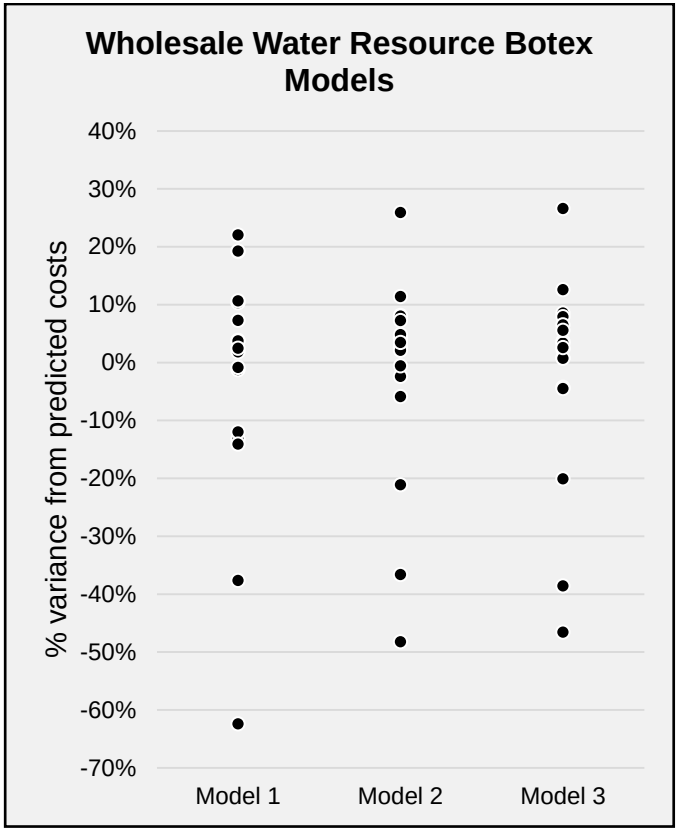
The models and corresponding coefficients presented in this pro forma are based on wholesale cost information for 17 companies[1]. Regressions were run in reference to the Master Wholesale Cost data file dated 27th February 2017, reflecting the latest updates and amendments to the data.

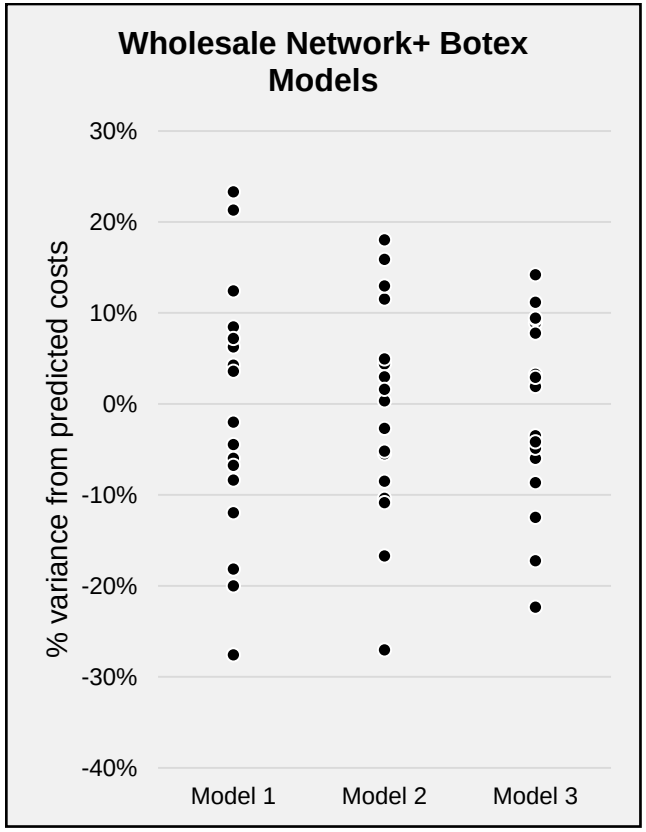
To note, the reported P-stats are based on the cluster robust standard error.

[1] Wholesale cost data for Bournemouth and South West Water have been appropriately combined

	Total Botex			Water Resources Botex			Network Plus Botex		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Dependent variable	Ln(total botex per connected property aggregate)	Ln(total botex per connected property aggregate)	Ln(total botex per connected property aggregate)	Ln(water resource botex per connected property resources)	Ln(water resource botex per connected property resources)	Ln(water resource botex per connected property resources)	Ln(network plus botex per connected property resources)	Ln(network plus botex per connected property resources)	Ln(network plus botex per connected property resources)
Ln(DI/connected property) Unit: Ml/d per '000 connected property	0.718 (0.116)	0.834** (0.0160)	0.753* (0.0554)				0.793* (0.0632)	1.050** (0.0239)	0.973*** (0.00295)
Ln(length of mains/ connected property) Unit: km per '000 connected property	0.279 (0.231)	0.346* (0.0972)	0.454*** (0.00885)				0.607*** (0.00789)	0.494** (0.0113)	0.461* (0.0755)
Ln(length of raw mains and conveyors/DI) Unit: km per Ml/d	0.0414 (0.624)				0.134* (0.0534)	0.109* (0.0778)			
Share of water treated at level 5 and above Unit: %	0.354** (0.0165)		0.193 (0.222)				0.257 (0.144)	0.143 (0.364)	
Length of mains laid pre-1940/Total length of mains Unit: %	0.270 (0.358)	0.987*** (0.00598)	0.439 (0.140)				0.805** (0.0120)	0.487 (0.152)	1.197*** (0.000425)
Length of renewed and relined mains/Total length of mains Unit: %	11.11 (0.278)	32.36** (0.0214)	17.71* (0.0902)				18.12* (0.0929)	15.64 (0.157)	34.12** (0.0108)
Ln(average pumping head aggregate) Unit: m.hd	0.228 (0.146)	0.0255 (0.806)	0.102 (0.559)						
Year15 Unit:dummy variable	-0.0153 (0.460)	0.0173 (0.583)	-0.00309 (0.892)	-0.0391* (0.0841)	-0.0385 (0.108)	-0.0381 (0.107)	0.000301 (0.990)	0.00268 (0.912)	0.0242 (0.468)
Year16 Unit:dummy variable	-0.00414 (0.911)	0.0623 (0.220)	0.0229 (0.590)	-0.0290 (0.569)	-0.0345 (0.498)	-0.0344 (0.492)	0.0243 (0.521)	0.0290 (0.432)	0.0719 (0.138)
Year17 Unit:dummy variable	-0.0122 (0.699)	0.0633* (0.0924)	0.0190 (0.583)	-0.0473 (0.448)	-0.0630 (0.305)	-0.0626 (0.297)	0.0244 (0.461)	0.0315 (0.301)	0.0766** (0.0458)
Surface water treated / Total water treated Unit: %		0.541** (0.0203)							0.526*** (0.00626)
Share of water from reservoirs Unit: %		0.0511 (0.753)	0.272** (0.0267)	0.680** (0.0256)	0.673** (0.0268)	0.624*** (0.00985)		0.206* (0.0592)	
Ln(number of sources / DI) Unit; sources per Ml/d		0.132 (0.172)		0.167** (0.0250)	-0.0338 (0.601)	0.000739 (0.993)			0.113 (0.325)
Share of water from boreholes Unit: %				-0.224 (0.383)	0.119 (0.657)				
Ln(average pumping head resources) Unit: m.hd					0.180* (0.0674)	0.171 (0.113)			
Ln(average pumping head network) Unit: m.hd								0.140 (0.294)	
Constant	-3.718*** (0.000960)	-3.131*** (0.000207)	-3.699*** (4.61e-05)	-4.006*** (0)	-5.013*** (0)	-4.872*** (3.27e-09)	-3.733*** (2.55e-05)	-3.881*** (2.31e-06)	-3.437*** (0.000629)
R2 adj ('R2 overall' in the case of random effects)	0.61 (R2)	0.73 (R2)	0.67 (R2)	0.45 (R2)	0.55 (R2)	0.54 (R2)	0.53 (R2)	0.65 (R2)	0.70 (R2)
Reset test	0.94	0.24	0.74	0.43	0.44	0.45	0.67	0.54	0.15
VIF	1.99	3.66	1.95	2.21	3.11	1.51	1.48	1.84	2.94
Method (eg OLS or RE)	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
N (sample size)	68	68	68	68	68	68	68	68	68







Negative numbers mean actual costs are lower than predicted (modelled) costs

Econometric model formula:

A1. $\ln(\textit{total retail operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{single service customers}_{it}) + \beta_2 \ln(\textit{dual service customers}_{it}) + \beta_3\, \textit{flats}_{it} + \beta_4\, \textit{IMD income}_{it} + \beta_5 \ln(\textit{average wholesale bill}_{it}) + \varepsilon_{it}$

A2. $\ln(\textit{bad debt related operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{single service customers}_{it}) + \beta_2 \ln(\textit{dual service customers}_{it}) + \beta_3\, \textit{IMD income}_{it} + \beta_4 \ln(\textit{average wholesale bill}_{it}) + \beta_5\, \textit{internal migration}_{it} + \varepsilon_{it}$

A3. $\ln(\text{non-}\textit{bad debt related operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{single service customers}_{it}) + \beta_2 \ln(\textit{dual service customers}_{it}) + \beta_3\, \textit{metered households}_{it} + \beta_4\, \textit{metered household density}_{it} + \beta_5 \ln(\text{peak traffic speed}_{it}) + \varepsilon_{it}$

A4. $\ln(\textit{total retail operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{single service customers}_{it}) + \beta_2 \ln(\textit{dual service customers}_{it}) + \beta_3\, \textit{metered households}_{it} + \beta_4\, \textit{metered household density}_{it} + \beta_5\, \textit{flats}_{it} + \beta_6 \ln(\textit{peak traffic speed}_{it}) + \beta_7\, \textit{IMD income}_{it} + \beta_8 \ln(\textit{average wholesale bill}_{it}) + \varepsilon_{it}$

A5. $\ln(\textit{total retail operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{single service customers}_{it}) + \beta_2 \ln(\textit{dual service customers}_{it}) + \beta_3\, \textit{IMD income}_{it} + \beta_4\, \textit{property repossessions}_{it} + \beta_5 \ln(\textit{average wholesale bill}_{it}) + u_i + v_{it}$

A6. $\ln(\textit{bad debt related operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{single service customers}_{it}) + \beta_2 \ln(\textit{dual service customers}_{it}) + \beta_3\, \textit{IMD income}_{it} + \beta_4 \ln(\textit{average wholesale bill}_{it}) + u_i + v_{it}$

A7. $\ln(\text{non-}\textit{bad debt related operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{single service customers}_{it}) + \beta_2 \ln(\textit{dual service customers}_{it}) + \beta_3\, \textit{metered households}_{it} + \beta_4 \ln(\text{peak traffic speed}_{it}) + \beta_5\, \textit{time trend}_t + u_i + v_{it}$

A8. $\ln(\textit{total retail operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{single service customers}_{it}) + \beta_2 \ln(\textit{dual service customers}_{it}) + \beta_3\, \textit{metered households}_{it} + \beta_4\, \textit{flats}_{it} + \beta_5\, \textit{IMD income}_{it} + \beta_6\, \textit{property repossessions}_{it} + \beta_7 \ln(\textit{average wholesale bill}_{it}) + u_i + v_{it}$

B1. $\ln(\textit{total retail operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{total customers}_{it}) + \beta_2\, \textit{IMD income}_{it} + \beta_3\, \textit{property repossessions}_{it} + \beta_4 \ln(\textit{average wholesale bill}_{it}) + \varepsilon_{it}$

B2. $\ln(\textit{bad debt related operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{total customers}_{it}) + \beta_2\, \textit{IMD income}_{it} + \beta_3 \ln(\textit{average wholesale bill}_{it}) + \varepsilon_{it}$

B3. $\ln(\text{non-}\textit{bad debt related operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{total customers}_{it}) + \beta_2 \ln(\textit{single service customers}_{it}) + \beta_3\, \textit{metered households}_{it} + \beta_4 \ln(\text{peak traffic speed}_{it}) + \varepsilon_{it}$

B4. $\ln(\textit{total retail operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{total customers}_{it}) + \beta_2 \ln(\textit{single service customers}_{it}) + \beta_3\ \textit{metered properties}_{it} + \beta_4\ \textit{IMD income}_{it} + \beta_5\ \textit{property repossessions}_{it} + \beta_6 \ln(\textit{average wholesale bill}_{it}) + \varepsilon_{it}$

B5. $\ln(\textit{total retail operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{total customers}_{it}) + \beta_2 \ln(\textit{single service customers}_{it}) + \beta_3\ \textit{property repossessions}_{it} + \beta_4 \ln(\textit{average wholesale bill}_{it}) + u_i + v_{it}$

B6. $\ln(\textit{bad debt related operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{total customers}_{it}) + \beta_2\ \textit{IMD income}_{it} + \beta_3\ \textit{property repossessions}_{it} + \beta_4 \ln(\textit{average wholesale bill}_{it}) + u_i + v_{it}$

B7. $\ln(\textit{non-bad debt related operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{total customers}_{it}) + \beta_2 \ln(\textit{single service customers}_{it}) + \beta_3\ \textit{metered households}_{it} + \beta_4 \ln(\text{peak traffic speed}_{it}) + \beta_5\ \text{time trend}_t + u_i + v_{it}$

B8. B8. $\ln(\textit{total retail operating costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{total customers}_{it}) + \beta_2 \ln(\textit{single service customers}_{it}) + \beta_3\ \textit{metered households}_{it} + \beta_4\ \textit{property repossessions}_{it} + \beta_5 \ln(\textit{average wholesale bill}_{it}) + u_i + v_{it}$

Description of dependent variable

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.

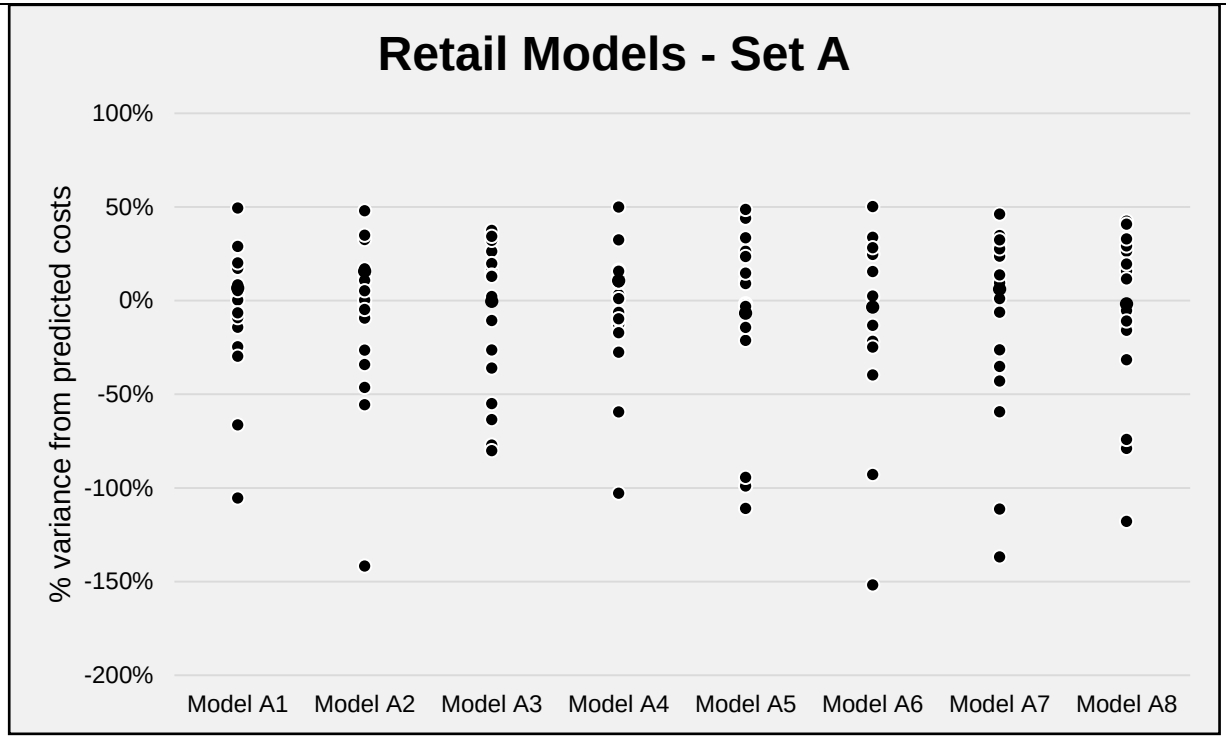
Overall, we consider the models across the suite to be valid. We regard pooled OLS and random effects GLS models as complementary: random effects models can distinguish between noise and other components of the error term, but unlike pooled OLS, produce time-invariant inefficiency estimates. Further, 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.

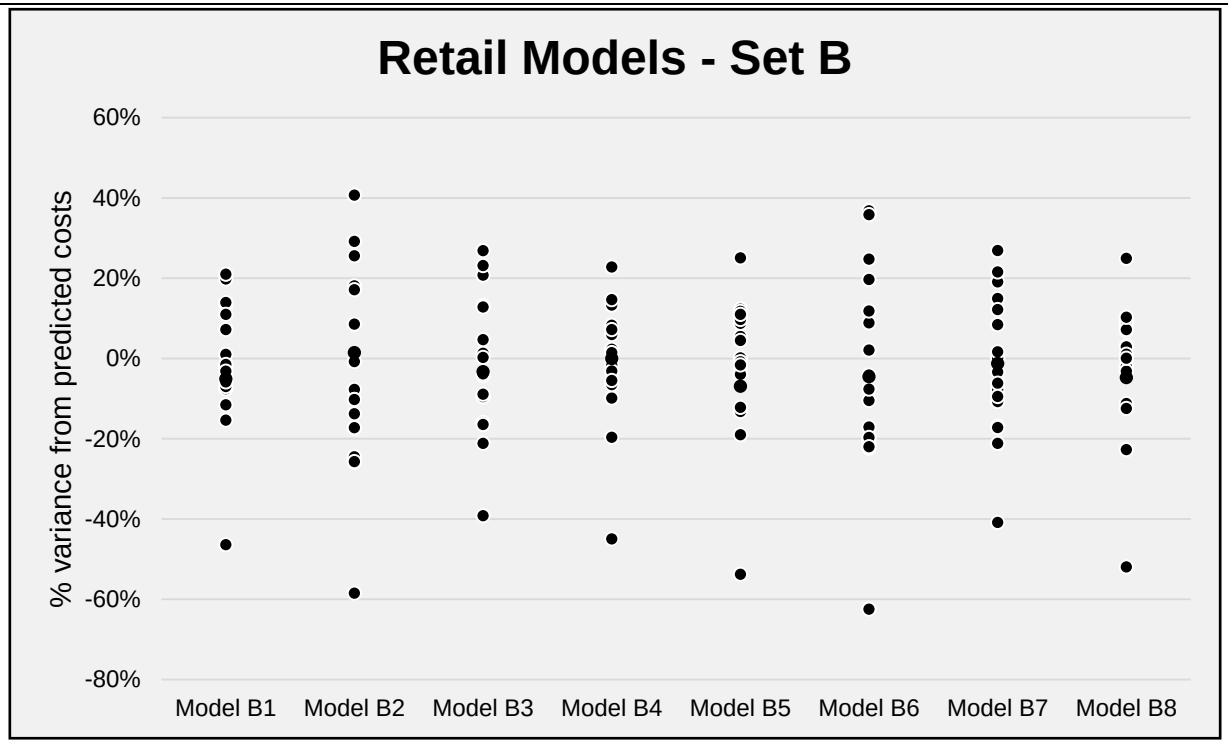
Data Sources for Variables used:

Variable name / measure	Cost driver	Source	Years available
Number of single service customers	Scale and scope factors	Company data share	2011/12-2016/17
Number of dual service customers	Scale and scope factors	Company data share	2011/12-2016/17
Proportion of metered households	Meter penetration	Company data share	2011/12-2016/17
Number of metered households relative to mains length (km)	Density	Company data share	2011/12-2016/17
Flats as proportion of households	Ease of meter reads	ONS	2011
Average peak hour traffic speed on A roads	Congestion	DfT	2011-2016
IMD income score	Regional socioeconomic performance	ONS	2010,2015
Mortgage repossessions relative to population	Regional socioeconomic performance	ONS	2011/12-2015/16
Wholesale bill size	Wholesale bill	Company data share	2011/12-2016/17
Population flow rate	Population transience	ONS	2011/12-2015/16

	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

Chart Percentage variance between modelled and actual costs	 Negative numbers mean actual costs are lower than predicted (modelled) costs

	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	 Negative numbers mean actual costs are lower than predicted (modelled) costs