March 2018

Economic Insight Ltd

HOUSEHOLD RETAIL COST ASSESSMENT FOR PR19

Final report for Bristol Water and Wessex Water

CONTENTS

1. Executive Summary 3

1.1 Context and aims 3

1.2 Our approach and summary of our descriptive statistics analysis 4

1.3 Econometric modelling summary 5

1.4 Summary of efficiency gaps 7

1.5 Key conclusions 8

1.6 Recommendations 9

2. Introduction 11

2.1 Background context to our work 11

2.2 The scope of our work 13

2.3 Key definitions and terms 15

2.4 Report structure 15

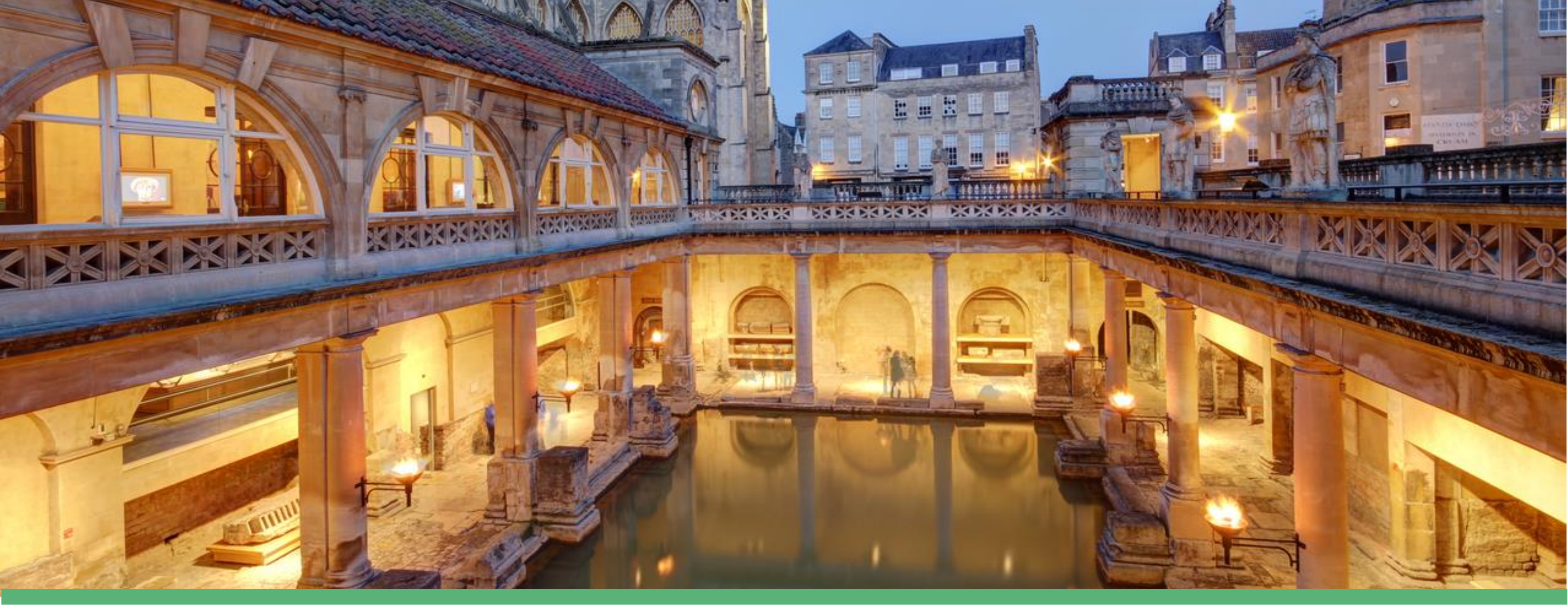
3. Key Issues, Cost Drivers and Descriptive Statistics 16

3.1 Our approach to considering retail cost assessment at PR19 16

3.2 Summary analysis of aggregate household retail costs 18

3.3 Descriptive statistics of key cost drivers 20

4. Econometric Modelling and Results 54
4.1 Our approach 54
4.2 Dataset 60
4.3 General to specific modelling 62
4.4 Results 66
4.5 Summary diagnostic testing 70
5. Efficiency Gap Estimates 72
5.1 Issues in the calculation of efficiency gaps 72
5.2 Efficiency assumptions 75
5.3 Efficiency scores 76
5.4 Wider sector comparators 81
6. Conclusions and Recommendations 84
6.1 Conclusions 84
6.2 Recommendations 85
7. Appendix: Diagnostic Tests 87



1. Executive Summary

This report, on behalf of Bristol and Wessex Water, sets out an econometric cost assessment analysis in relation to the retail household control at PR19. Our work, which has benefitted from independent expert input from Drs Anthony and Karli Glass (Centre for Productivity and Performance, Loughborough) is supportive of econometric methods being a valid and practical approach to cost assessment for retail in the water sector. Our work also highlights that: (i) due to the historical focus on scale and bad debt related issues, there is a risk that other, valid, drivers may not have been considered in detail to date; (ii) relatedly, because scale effects tend to 'dominate' in statistical models, careful thought must be given as to the balance between statistical robustness and intuition as to the 'in principle' drivers of retail costs; (iii) consistent with Ofwat's draft proposals, there is no strong 'regional wage' dimension to retail costs; and (iv) when converting econometric benchmarking results to efficiency gaps (i.e. targets), care must be taken to consider assumptions holistically, to avoid setting unduly challenging (and implausible) efficiency challenges.

1.1 Context and aims

In its final methodology for PR19, Ofwat proposes to adopt an econometric modelling approach to assessing efficient costs for the household retail price control. Bristol Water (Bristol) and Wessex Water (Wessex) commissioned Economic Insight to provide an independent view as the appropriate approach to retail cost assessment at PR19. The main objectives of our work were:

- Firstly, to provide Bristol and Wessex with a better **understanding of their 'true' retail cost efficiency**, which could - in turn - inform their PR19 Business Plans for household retail.
- Secondly, to help **provide thought leadership in this important area**, so that the companies can contribute constructively to developing a robust and practical approach to retail cost assessment.

1.2 Our approach and summary of our descriptive statistics analysis

Our overall approach starts from 'first principles'; and we considered afresh what the key cost drivers for retail (which are outside of efficient management control) might be. This in part reflects the fact that, historically, the focus has predominantly been on scale and bad debt (and, more specifically, deprivation) in a retail context. Thus, it seemed to us, other potentially valid drivers may not have been considered in detail. Consequently, our work begins with a comprehensive descriptive statistics analysis.

Key points highlighted by this include:

- That, as has been well-established, retail is dominated by 'scale' effects – and data are **consistent with the presence of both 'economies of scale' and 'economies of scope'.**

- **In relation to bad debt**, and consistent with the evidence reviewed at PR14, we find that **both socioeconomic factors, or deprivation** (which might affect customers' propensity to go into arrears / default) **and average wholesale bill size** (which impacts the absolute value at risk through default) **are valid drivers.** There are numerous measures of socioeconomic performance; and our descriptive statistics analysis is generally consistent with a range of measures being plausible and credible. In addition, we consider that population transience (the propensity of people to move in to, or out of, a region) also affects debt costs. For example, it might be related to the propensity to fall into arrears, but also might positively impact company debt management costs. We further consider whether transience might impact retail costs more generally, as it might be associated with higher account management costs (e.g. companies needing to open, close, or transfer accounts, as customers move house). Our descriptive statistical analysis is more consistent with the latter effect than the former, although the reverse is true within our econometric analysis.

- Also consistent with the PR14 approach, we find **meter penetration to be positively associated with retail costs** (i.e. the greater the number, or proportion, of meters a company has, the higher its metering costs are).

- **A range of other factors may also impact metering costs,** such as the configuration of housing stock, metering (i.e. geographic) density; and congestion. These factors have not been considered in detail until now, but appear intuitively sound and are supported by data. Given that metering reflects only a small proportion of the overall retail cost stack, however, it is not clear a priori, whether such measures will perform well in statistical models. This goes to a wider issue regarding the **balance between statistical robustness and intuition.**

- **Our descriptive statistics are not consistent with regional wages being an important cost driver for retail household.** This is true both in relation to overall regional wages, and specific occupation codes most relevant to retail. This can intuitively be explained by the fact that the majority of labour-intensive water retail activities do not have an inherently regional dimension (many activities can be outsourced geographically). This is also consistent with Ofwat's position.

- **While, in principle, economics tells us that there should be a cost / quality relationship, our analysis is unable to identify one.** There could be multiple explanations for this and, of course, our finding is in part a function of the available data which – in relation to quality – primarily consists of the Service Incentive Mechanism (SIM). However, the fact that a relationship *for the specific measures available* cannot be identified does not imply that cost assessment should be undertaken independently of a consideration of quality performance. This is an area that merits further consideration.

Our descriptive statistical analysis highlights three key issues:

- Firstly, that because scale and bad debt related issues are likely to be the main drivers of retail costs overall, there is a risk that these mask other intuitively valid drivers. Therefore, when first exploring the data, care must be taken to precisely identify potential factors; and a hypothesis for how they might impact costs. This may also suggest that, in an econometric modelling context, there will be some intuitively valid drivers that may not be statistically significant. Consequently, in relation to retail, the question of how best to balance significance and intuition is likely to be particularly pertinent.
- Secondly, when one considers the drivers at a more granular level, some may be related, yet could drive costs in different directions. For example, the issues of 'traffic congestion' and 'housing stock' might broadly relate to issues of urbanisation or rurality. Here, one might expect increased congestion to generally put upward pressure on retail costs. However, this might be positively associated with areas having a higher proportion of flats, which in turn might mitigate retail costs (if, for example, meter reading is less costly in relation to flats).
- Thirdly, of course, descriptive statistics alone will not necessarily reveal the nature of cost and driver relationships. For example, while intuitively plausible, our descriptive analysis does not find a relationship between population transience and debt costs. However, once the 'larger' debt cost drivers are controlled for statistically, a robust relationship does emerge.

1.3 Econometric modelling summary

Our econometric modelling has the following key features:

- A **combination of top-down models** for total retail operating costs, and **bottom-up models** for bad debt and non-bad debt related retail operating costs.[1]

- A **general to specific approach**, which incorporates a relatively 'liberal' approach to statistical significance by retaining variables that are significant at levels approaching 10%.

- **Alternative econometric models** that balance statistical significance with engineering intuition by including variables that are correctly signed, but which would not be included within a strict general to specific approach.

[1] *As we explain in the main body of our report, these terms align to Ofwat's definitions of retail cost categories for cost assessment at PR19 (i.e. by 'bad debt related we mean doubtful debt and debt management).*

- A combination of **pooled models** estimated using Ordinary Least Squares (OLS) and **random effects models**, estimated using Generalised Least Squares (GLS).
- Two approaches to the **incorporation of customer numbers and scope** (dual versus single bill customers). Model set A includes separate dual and single service customer variables. Model set B includes separate variables for the total number of customers and the number of single service customers.

This generated a suite of 16 models, as summarised in the table below.

Table 1: Suite of econometric cost models

Model	Dependent variable	Panel structure	Estimation technique	General to specific approach	Approach to number of customers
A1	Total retail operating costs	Pooled	Ordinary Least Squares	Statistical significance	Separate dual and single service customer variables
A2	Bad debt related retail operating costs	Pooled	Ordinary Least Squares	Statistical significance	Separate dual and single service customer variables
A3	Non-bad debt related retail operating costs	Pooled	Ordinary Least Squares	Statistical significance	Separate dual and single service customer variables
A4	Total retail operating costs	Pooled	Ordinary Least Squares	Alternative approach	Separate dual and single service customer variables
A5	Total retail operating costs	Random effects	Generalised Least Squares	Statistical significance	Separate dual and single service customer variables
A6	Bad debt related retail operating costs	Random effects	Generalised Least Squares	Statistical significance	Separate dual and single service customer variables
A7	Non-bad debt related retail operating costs	Random effects	Generalised Least Squares	Statistical significance	Separate dual and single service customer variables
A8	Total retail operating costs	Random effects	Generalised Least Squares	Alternative approach	Separate dual and single service customer variables
B1	Total retail operating costs	Pooled	Ordinary Least Squares	Statistical significance	Total customers; single service customers
B2	Bad debt related retail operating costs	Pooled	Ordinary Least Squares	Statistical significance	Total customers; single service customers
B3	Non-bad debt related retail operating costs	Pooled	Ordinary Least Squares	Statistical significance	Total customers; single service customers
B4	Total retail operating costs	Pooled	Ordinary Least Squares	Alternative approach	Total customers; single service customers
B5	Total retail operating costs	Random effects	Generalised Least Squares	Statistical significance	Total customers; single service customers
B6	Bad debt related retail operating costs	Random effects	Generalised Least Squares	Statistical significance	Total customers; single service customers
B7	Non-bad debt related retail operating costs	Random effects	Generalised Least Squares	Statistical significance	Total customers; single service customers
B8	Total retail operating costs	Random effects	Generalised Least Squares	Alternative approach	Total customers; single service customers

Source: Economic Insight

1.4 Summary of efficiency gaps

To convert the results of econometric benchmarking into efficiency gaps (i.e. percentage cost savings targets), regulators need to address two key difficulties:

- Efficiency scores are estimated using model residuals, but it is uncertain how much of any model residual represents inefficiency and how much represents other factors, including random noise and model specification problems.

- It is uncertain how much of any efficiency gap can be closed in practice, and how quickly this can be done.

Further, when drawing on evidence from a suite of models, they also need to decide the weight to attach to different models which, inevitably, generate different results.

A range of 'policy tools' are available to address these issues. Uncertainty over inefficiency's contribution to residuals can be addressed by some combination of: (i) residual adjustments; and (ii) the use of benchmarks that are 'less demanding' than the absolute frontier (i.e. the minimum residual), such as using the average, upper quartile or upper quintile, as the benchmark. There is no clear 'right answer', although one can strictly say that it is inappropriate to use the unadjusted minimum residual (i.e. absolute frontier). Uncertainty over the practicality and speed of closing efficiency gaps can be addressed by making downward percentage adjustments efficiency gap estimates; and/or by allowing firms a glide path to efficient costs.

For benchmarking PR19 household retail, we have developed three scenarios with varying assumptions relating to the above issues. This provides a transparent basis for showing how differences in these assumptions can affect the implied efficiency challenge. This, in turn, allows stakeholders to consider which assumptions / scenario they consider to be most appropriate. These scenarios are set out in the table below.

Table 2: Underlying assumptions for efficiency gap calculation

Parameter	Low case	Central case	High case
Model weights	Equal weights	Equal weights	Equal weights
Residual adjustment	None	None	None
Benchmark	Average	Upper quartile	Upper quintile
Glide path	Five-year	None	None

Source: Economic Insight

The implied total efficiency gaps across these three scenarios are set out in the following table. As our 'central' and 'high case' scenarios assume no glide path, under these approaches it is assumed that the entirety of the total efficiency gap would be closed in 'year 1' of PR19. In contrast, in our 'low case' a glide path is assumed, and so the total efficiency gap would be spread over the 5 years of PR19.

Table 3: Total efficiency challenges (i.e. % gaps)

Company	Low			Central			High		
	Model set A	*Model set B*	*Average*	*Model set A*	*Model set B*	*Average*	*Model set A*	*Model set B*	*Average*
AFW	42.5%	6.4%	24.5%	53.1%	17.6%	35.3%	55.2%	18.9%	37.1%
ANH	0.0%	0.0%	0.0%	17.5%	5.8%	11.6%	21.2%	7.3%	14.2%
BRL	**0.0%**	**0.0%**	**0.0%**	**10.8%**	**0.0%**	**5.4%**	**14.8%**	**1.3%**	**8.0%**
DVW	0.0%	5.4%	2.7%	0.0%	16.5%	8.2%	0.0%	17.8%	8.9%
NES	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.9%	0.4%
PRT	0.0%	0.0%	0.0%	3.4%	3.4%	3.4%	8.3%	5.1%	6.7%
SES	0.0%	0.0%	0.0%	0.0%	10.7%	5.4%	0.0%	12.2%	6.1%
SEW	40.2%	0.0%	20.1%	51.8%	9.2%	30.5%	54.0%	10.7%	32.3%
SRN	19.8%	36.4%	28.1%	35.0%	44.3%	39.7%	38.0%	45.2%	41.6%
SSC	11.9%	5.0%	8.4%	28.7%	16.4%	22.6%	32.0%	17.8%	24.9%
SVT	1.2%	0.0%	0.6%	20.4%	5.1%	12.8%	24.0%	6.6%	15.3%
SWT	1.1%	4.9%	3.0%	20.0%	16.9%	18.4%	23.5%	18.2%	20.9%
TMS	0.0%	7.0%	3.5%	11.4%	18.5%	15.0%	15.5%	19.8%	17.7%
UU	45.2%	5.0%	25.1%	55.5%	17.1%	36.3%	57.5%	18.4%	38.0%
WSH	6.1%	17.3%	11.7%	23.3%	28.0%	25.6%	26.7%	29.1%	27.9%
WSX	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**5.6%**	**2.8%**	**0.0%**	**7.1%**	**3.5%**
YKY	0.0%	0.0%	0.0%	11.5%	0.0%	5.7%	15.5%	0.0%	7.8%

Source: Economic Insight

1.5 Key conclusions

Our main conclusions arising from the analysis set out in this report are as follows:

- **It is possible to identify econometric benchmarking models for household retail that perform well on measures of statistical robustness and are intuitively sound.** Given this, the analysis contained here, collectively, is supportive of the use of econometric modelling for setting efficient household retail costs at PR19.

- Related to the above**, our modelling identifies a range of key cost drivers that are reasonably outside of efficient management control,** for inclusion in benchmarking. In particular, key cost drivers include:
  - measures of single and dual serve customers (to identify economies of scale and scope);
  - meter penetration;
  - socioeconomic factors e.g. Index of Multiple Deprivation (IMD); and
  - average wholesale bills size.

- **Regional wages are not found to be either an intuitively sound, nor a statistically valid, driver of household retail costs.** Consequently, we do not think there is a case for them being included within any econometric cost assessment model.

- **Because of the predominance of 'scale' and 'bad debt' related cost drivers – there is a risk of overlooking other, intuitively sensible, drivers of cost.** Indeed, our descriptive statistics and modelling analysis identified several factors, which hitherto have received relatively little attention in a retail cost assessment context. These include, for example:
  - density of metered properties;
  - congestion;
  - housing stock; and
  - population transience.

- **The implied efficiency challenges may be overly demanding, depending on the assumptions made.** Because there are several 'steps' that must be taken to convert econometric modelling into efficiency gaps, there is a risk that, if each is considered individually, and 'aggressive' assumptions are made, the implied results are unduly demanding (and indeed, may be considered implausible).

- **The way in which scale and scope are accounted for within the econometric models can have a significant impact on implied efficiency scores, in some cases.** While some companies have similar scores under both of the approaches to measuring customer numbers incorporated within our suite of models, for some companies implied efficiency gaps vary more materially across the two methods. As we explain in the main body of our report, as both approaches have analytical merit, it is important to be mindful of this.

- **It is important not to conflate benchmarking with setting a future profile of allowed costs over time –** where in the latter, one may very well wish to take a case of 'foreseeable' changes that ultimately will impact retail costs. For example, wholesale bill size is 'out of efficient retail management control' and therefore, should be controlled for when undertaking benchmarking analysis. However, when setting a forward view of allowed costs, it would seem to be legitimate for Ofwat to consider known changes in wholesale bill size – reflecting, for example: (i) the wholesale efficiency challenge Ofwat sets (which, all else equal, will *reduce* required retail costs); and (ii) general inflation allowed for at the wholesale level (which, all else equal, will *increase* required retail costs).

1.6 Recommendations

Following from the above, our recommendations are as follows:

- Consistent with Ofwat's draft proposals, **an econometric approach should be used to set allowed costs for household retail at PR19.**

- **We recommend that Ofwat pays particular attention to the wider range of potentially valid explanatory variables** (outside of efficient management control) which might be 'crowded out' by the predominance of 'scale' and 'bad debt' related drivers. We further suggest that the precise way in which such factors might impact costs should be evaluated with care – as the historical focus on issues such as bad debt and deprivation has meant these issues have not been considered in detail to date.

- **When converting econometric results to efficiency challenges, the assumptions should be considered holistically** and the resultant efficiency gaps 'sense checked', to ensure they are plausible and defendable.

- **Care needs to be taken when defining the frontier**; and there are dangers in being overly reliant on the performance of any one company. This does not necessarily mean that, for example, upper quartile performance is the 'right' answer. However, care and consideration should be given to the sensitivity of results to different definitions of the frontier.

- **Service quality should be a consideration when setting the frontier.** Our analysis found significant difficulties in incorporating service quality within cost models. For example, raw data suggest a negative correlation, while none of the models we tested including quality variables produced signs that accorded with our priors. As a matter of principle, however, at the frontier there must be a cost-quality trade-off. Therefore, our view is that this issue is best addressed at PR19 by the regulator paying attention to the quality performance of any firms identified as being a candidate for the 'frontier', to avoid setting unduly challenging or unduly lenient, targets.

2. Introduction

This chapter introduces our household retail cost assessment study, undertaken on behalf of Bristol and Wessex Water. We firstly set out the relevant background context to the work – where the key points to note include Ofwat's decision to apply an econometric approach to retail cost assessment at PR19; and the regulator's further guidance that it will examine two approaches: (i) one modelling 'total retail costs' and (ii) another modelling 'bad debt' and 'non-bad debt' costs separately. We then set out the overall scope of our work – where our main objectives are: (a) to assist Bristol and Wessex in genuinely understanding their 'true' efficiency performance for business planning purposes; and then (b) to help constructively contribute to the debate regarding the most appropriate approach to setting retail costs at PR19.

2.1 Background context to our work

At PR14, allowed revenues for the household retail control consisted of: (i) a cost allowance, which included an efficiency challenge based on the industry average cost to serve (ACTS); and (ii) an allowed net retail margin (set on a % EBIT basis). No automatic pass through of general inflation was allowed for.

In practice, Ofwat did not apply a 'pure' ACTS approach to set 'efficient' retail costs. Rather, reflecting a range of views from industry regarding factors that could impact efficient costs that would not be captured in a simple unit cost approach, Ofwat applied various 'adjustments' to the ACTS. These included:

- an economy of scope adjustment, relating to inherent efficiencies in serving both water and wastewater customers (Ofwat applied an adjustment factor of 1.3); and
- adjustments for metering levels (Ofwat calculated cost to serve separately for unmetered and metered customers, making adjustments based on the incremental cost of serving metered customers).

In addition to the above, companies could make 'special factor' cost claims relating to factors that could drive their retail costs, but were not captured by Ofwat's adjusted ACTS approach (Ofwat then assessed these claims against a range of criteria). In practice, two of the most material special factor claims related to:

- input price pressure, which amounted to >£37m of additional revenues across the industry; and
- deprivation's impact on bad debt (specifically, the combination of doubtful debt and debt management), which amounted to >£190m across the industry.

Other allowances included: pension deficit repair costs; allowances for 'new costs' to support investments; and a small number of company-specific issues.

Regarding the net retail EBIT margin, Ofwat set this at 1.0% for household customers. The primary source of evidence Ofwat relied upon was a report by PWC[2], which drew on both a comparative and return on capital approach. More weight was placed on the former (particularly regulatory precedent).

In July 2017, Ofwat published its draft PR19 methodology for consultation. Within this, the regulator set out the following proposals in relation to the household retail control.

- The retail control will continue to be set on a (weighted) average revenue basis.
- The duration of the control will be shortened to three years.
- Ofwat intends to **set efficient retail costs by using econometric benchmarking** (instead of using ACTS). Ofwat further stated that*: "if we are unable to produce robust econometric models, we propose to use an efficient cost to serve"*[3] (by which Ofwat explains they mean an 'average' unit cost approach, rather than econometrics, but with a more stringent benchmark).
- Ofwat is specifically proposing to develop both: (i) **an overall industry level model,** capturing all household retail operating costs; and (ii) **an approach where bad debt and non-bad debt related operating costs are modelled separately.**
- Ofwat's dependant variable will be retail operating costs – which includes both opex and capital related costs (specifically, depreciation).
- Ofwat specifically indicated that **regional wage costs will not be controlled for** in its modelling: *"We do not propose to account for variation in regional labour costs in our benchmarking analysis. With the exception of metering, which is around 5% of total retail costs, we consider that the impact of regional labour costs can be substantially mitigated or removed in retail activities."*[4]

'We do not propose to account for variation in regional labour costs in our benchmarking analysis.' Ofwat

- **There will be no glide path down to the efficient level of retail costs.** Ofwat's stated rationale for this is that: *"By 2020, companies will have had five years of the residential retail controls (introduced at PR14) to catch up to the efficient level of*

[2] *'Water retail net margins: A report prepared for Ofwat.' PWC (2014).*
[3] *'Delivering Water 2020: consultation on PR19 methodology Appendix 12: Securing cost efficiency.' Ofwat (2017); page 14.*
[4] *'Delivering Water 2020: Consulting on our methodology for the 2019 price review.' Ofwat (2017); page 184.*

retail costs."[5] It should be noted, however, that this does not mean that Ofwat's cost allowances for household retail will not have a 'profile' over PR19. For example, projected retail costs could move over time due to frontier shift; and / or because of underlying input price inflation (see later).

- Ofwat has stated that, to set an aggressive frontier, it will look at cost to serve **evidence from other sectors.**

- As per PR14, Ofwat will not allow for any automatic indexation for inflation within the retail control. However, the regulator has stated that: *"if appropriate, our efficient cost baselines will include an allowance for input price pressure."*[6]

- The 'return' element of household retail will continue to be set on a net EBIT margin basis.

In its final methodology, Ofwat largely confirmed its draft approach to setting the household retail control. However, the regulator did: (i) amend its position on the proposed length of control **from three, to five, years**; (ii) provided some further guidance on the **explanatory variables** it would explore in cost assessment, listing: the impact of single vs dual customers; metering levels; deprivation; and bill sizes; (iii) indicated that it would set the benchmark relative to *'efficient companies'.*[7]

2.2 The scope of our work

In the above context, Bristol and Wessex commissioned us to provide an independent view as to the appropriate approach to retail cost assessment at PR19. The main objectives of our work were to:

» Firstly, provide Bristol and Wessex with a better **understanding of their 'true' retail cost efficiency**, which could in turn inform their PR19 Business Plans for household retail.

» Secondly, help **provide thought leadership in this important area**, so that the companies can contribute constructively to developing a robust and practical approach to retail cost assessment.

Related to the above, the scope of our work included:

- **Developing household retail econometric model(s)** – including ultimately providing our views and advice as to which model(s) should be preferred for efficiency benchmarking at PR19.

- **Setting out our views as to the efficiency challenges implied by the models** (e.g. converting modelling results into efficiency gaps – addressing issues such as what 'weight' to attach to which models, if multiple variants are developed).

- **Our modelling work is consistent with Ofwat's overarching approach.** The scope of our work includes developing both: (i) overall retail operating cost

[5] *'Delivering Water 2020: Consulting on our methodology for the 2019 price review.' Ofwat (2017); page 186.*

[6] *'Delivering Water 2020: consultation on PR19 methodology Appendix 12: Securing cost efficiency.' Ofwat (2017); page 18.*

[7] *'Delivering Water 2020: Our final methodology for the 2019 price review Appendix 11: Securing cost efficiency.' Ofwat (2017).*

models; and (ii) separate bad debt and non-bad debt related efficiency benchmarking models. Similarly, our dependent variable is retail operating costs (i.e. opex and capital costs).

Not within the scope of this work is any analysis or consideration of how retail costs might be 'profiled' over PR19 for reasons unconnected to efficiency benchmarking. For example, to reflect: (i) frontier productivity gains; (ii) underlying input price inflation; and / or (iii) changes to underlying cost drivers over time (e.g. decreases to bad debt costs due to forecast improvements in the wider UK economy) which could be captured in separate statistical cost forecasting models.

DRs ANTHONY AND KARLI GLASS, AT THE CENTRE FOR PRODUCTIVITY AND PERFORMANCE AT LOUGHBOROUGH, PROVIDED EXPERT ACADEMIC INPUT INTO OUR WORK.

Our work has also benefitted from the expert independent input of Drs Anthony and Karli Glass, at the Centre for Productivity and Performance (University of Loughborough) who are known experts in the field of efficiency benchmarking. The scope of their input to our work included:

- Providing advice, in advance of our modelling, on a range of issues that we subsequently reflected in our overarching approach. Issues on which they advised included: criteria for general to specific modelling; weighting of models; parameter stability; model specification; cost / quality relationships; and the balance between statistical validity and intuition.

- Providing an independent review and critique of our modelling, which ultimately, we reflected upon in a number of refinements to our analysis. In commenting on our finalised modelling and results, they stated: *"Our overall assessment is that the approach to the modelling was appropriate and the estimation results were clearly reported and well interpreted…. The path Economic Insight followed when moving from the general models to the specific models was entirely reasonable, as they omitted variables with coefficients that had counterintuitive signs and those with coefficients that were not significant at the 10% level."*

2.3 Key definitions and terms

Throughout this report, we make use of various terms relating to cost measures. For clarity and consistency, key terms are set out in the table below.

Table 4: Definition of key terms

Measure	Definition
(1) Retail operating costs	The totality of household retail costs, including opex and capital costs: customer services; debt management; doubtful debts; meter reading; developer services; other opex; local authority rates; exceptional items; third party services; depreciation; and amortisation).
(2) Bad debt related retail operating costs	Refers to a subset of (1) – specifically: debt management and doubtful debt.
(3) Non-bad debt related retail operating costs	Refers to the subset of (1) not included in (2). Specifically, all household retail operating costs other than 'debt management'; and 'doubtful debt'.
(4) Single serve customers	Customers receiving only one of 'water services' or 'wastewater services' from the company in question.
(5) Dual serve customers	Customers receiving both 'water services' or 'wastewater services' from the company in question.
(6) Unique customers	The sum of (4) and (5) above. Note, at PR14 Ofwat adjusted this measure by a factor of 1.3 to reflect economies of scope.[8] In our report, no off-model adjustments are made.

Source: Economic Insight

2.4 Report structure

The remainder of this report is structured as follows:

- **Chapter 3** sets out the key issues relating to household retail cost drivers – and a summary of our descriptive statistical analysis.
- **Chapter 4** contains the results of our econometric benchmarking analysis.
- **Chapter 5** sets out the implied efficiency gaps.
- In **Chapter 6** we put forward our conclusions and recommendations.

The appendix provides more detailed model diagnostics.

[8] *'Setting price controls for 2015-20 – final methodology and expectations for companies' business plans.' Ofwat (July 2013).*

3. Key Issues, Cost Drivers and Descriptive Statistics

This chapter provides an overview of household retail costs including: (i) a discussion of our overarching approach to retail cost assessment, highlighting some key issues; (ii) a summary of the overall makeup of, and trends in, retail costs across the industry; and (iii) summary descriptive statistics analysis, which informed our approach to the econometric modelling, set out in the next chapter of our report.

3.1 Our approach to considering retail cost assessment at PR19

Before summarising our initial descriptive statistical analysis, it is worth setting out our overall approach to considering household retail cost assessment at PR19 – and relatedly, highlighting key issues to be mindful of.

- **We have taken a first principles approach.** Historically there has been considerable debate regarding retail cost factors. This has tended to be focused on a number of narrow issues – most obviously, bad debt related cost drivers, such as deprivation measures. This is for understandable reasons and reflects where value is concentrated in retail activities. However, given Ofwat's intention to adopt an econometric approach at PR19, we think there is merit in 'stepping back' from the historical areas of focus and thinking afresh about what really drives underlying retail costs. Accordingly, in our descriptive statistics analysis, we consider several factors not previously examined in any detail – where our intention is to approach questions with an open mind, rather than any strong presumption of validity.

- **Factors that ultimately merit inclusion in modelling should be both: (i) drivers of retail costs; and (ii) reasonably beyond efficient management control.** These principles are well-understood, and so we do not expand on them in any detail here. In general terms, the first point is often more straightforward to establish objectively, whereas the latter can sometimes be a matter of degree.

- **Related to the above, it is important to distinguish between: (i) retail efficiency benchmarking and (ii) the setting of allowed retail totex over time.** Our work is concerned with the former issue, rather than the latter. This matters because there may be certain factors which are outside of efficient retail management control – and so might need to be 'controlled for' in benchmarking; but nonetheless will change in a (to some degree) *predictable* way over time, meaning that Ofwat may wish to take that into account when setting allowed retail costs. The most obvious example of this is average wholesale bill size, which as we subsequently demonstrate (and has been previously accepted) impacts the debt elements of retail costs in particular. While, we suggest, this is outside of efficient retail management control for benchmarking purposes, clearly Ofwat will 'know' the profile of wholesale bills for PR19 – which will reflect, amongst other things, deductions for efficiency and indexation for general inflation. Ofwat may, therefore, wish to reflect such profiling in its retail cost allowances over time. However, it is important not to conflate this with benchmarking (and therefore, retail management control).

RETAIL IS DOMINATED BY 'SCALE'. THIS BOTH MAKES IT EASY TO OVERLOOK POTENTIALLY VALID COST DRIVERS, BUT ALSO ULTIMATELY RAISES THE ISSUE OF HOW BEST TO BALANCE 'STATISTICAL SIGNIFICANCE' AGAINST 'ENGINEERING INTUITION' IN ANY MODELLING FRAMEWORK.

- **Retail is dominated by scale.** Various previous analyses have repeatedly shown that, above all else, retail costs can predominantly be explained by customer numbers. Importantly, this issue has the potential to cloud the interpretation of other descriptive analysis because, without controlling for scale, it is easy to overlook intuitively valid drivers of retail costs. Given this, when undertaking initial analyses of potential drivers and their measures, it is important to think carefully about exactly "how" they might impact retail costs, and therefore, which combinations of 'cost measure' and 'driver measure' are most pertinent to focus on.

- **Following from the above, the issue of balancing statistical significance and intuition is particularly important.** In any statistical cost efficiency benchmarking analysis, it is important to be mindful of the balance between 'statistical validity' and 'engineering intuition'. The predominance of 'scale' (and bad debt factors, such as deprivation) as a cost driver in retail makes this issue particularly important. Specifically, when retail costs are considered at a more granular level (and as shown in our analysis subsequently) one can identify a number of intuitively sensible drivers. However, when put within the context of a broader statistical modelling approach, in which certain other factors 'dominate', many of these *may* not be statistically significant. Accordingly, the question of how to balance statistical fit and intuition seems especially pertinent here.

3.2 Summary analysis of aggregate household retail costs

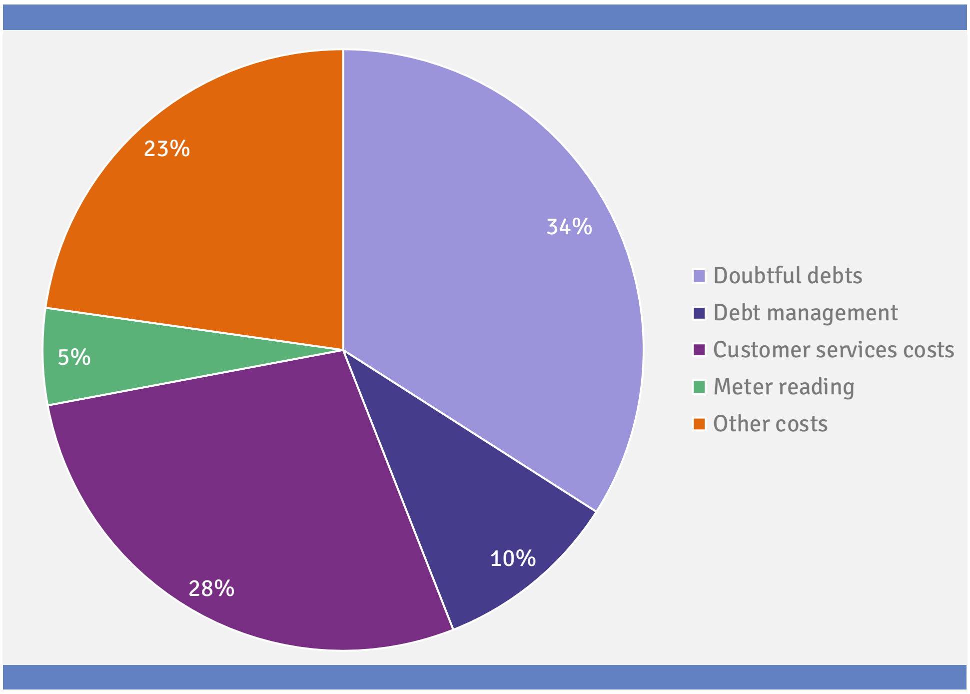
Before setting out our descriptive statistics analysis in relation to potential household retail cost drivers, it is useful to firstly examine the makeup of retail related costs and changes to them over time.

3.2.1 The composition of household retail costs

The following figure shows the makeup of retail costs for the industry as whole, in 2016/17. As can be seen – key cost categories include:

- doubtful debts 34%;
- debt management 10%;
- customer services costs 28%; and
- meter reading 5%.

Figure 1: Split of industry household retail costs



Source: Economic Insight

Here, the main issues to be mindful of are:

- Firstly, that doubtful debt and debt management costs are closely interrelated – in the sense that companies typically have to optimise the two collectively, as investment in improved debt management can lead to reduced doubtful (and bad) debt. As such, any approach to cost assessment should ideally ensure that these issues are evaluated concurrently.

- Secondly, the drivers of doubtful debt and debt management costs are intrinsically different to those for other retail functions (which are more 'service provision' in nature). This provides a rationale for considering cost assessment for these items (combined, as noted above) separately from the remainder of retail costs. This is consistent with Ofwat's proposed approach at PR19 – but also consistent with PR14, to the extent that there were special factor cost claims that focused specifically on the debt issue.

- Thirdly, and related to the above, we would further suggest that the key drivers of metering related costs are also relatively distinct (this is explained further on the following summary of our descriptive statistics). This matters because metering costs are a relatively small proportion of the total. Consequently, factors which intuitively should affect metering costs, and which can be shown with data and evidence to affect them, may not prove to be statistically significant within 'wider' retail cost models.

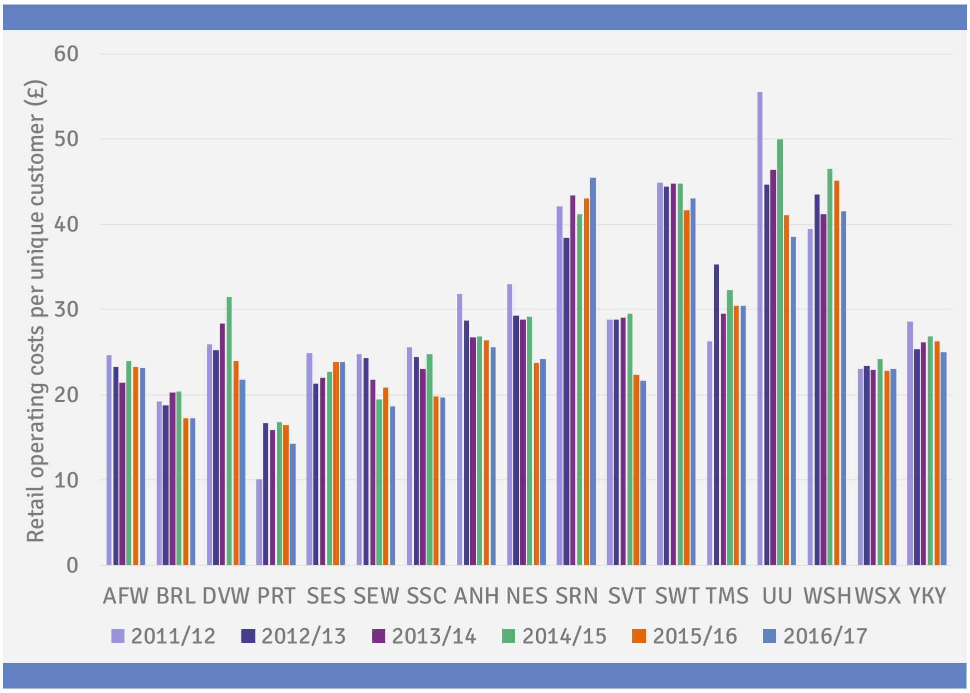
Arguably, the above could point to modelling metering costs separately. However, we also recognise the need to be proportionate in any approach to cost assessment – and the scope of our work is focused on the broader framework already set down by Ofwat.

3.2.2 Retail costs over time

THERE IS NO OBVIOUS PATTERN TO CHANGES IN RETAIL COSTS OVER TIME.

It is also helpful to understand how household retail costs have changed over time. The following figure therefore shows retail operating cost per unique customer, by company, over time from the financial year 2011/12 to 2016/17.

Figure 2: Retail operating costs per unique customer over time



Source: Economic Insight

In relation to the above, our main observations are that:

- There is no obvious 'pattern' or trend in average retail costs over time across the companies. In some cases, retail costs go up over time, in others they fall, and for many there is no clear trend one way or another.

- For some companies, particularly UU, figures for 2011/12 appear to be materially different to other years. This raises legitimate questions as to whether this reflects 'actual' cost differences, or issues around cost allocation and recording.

3.3 Descriptive statistics of key cost drivers

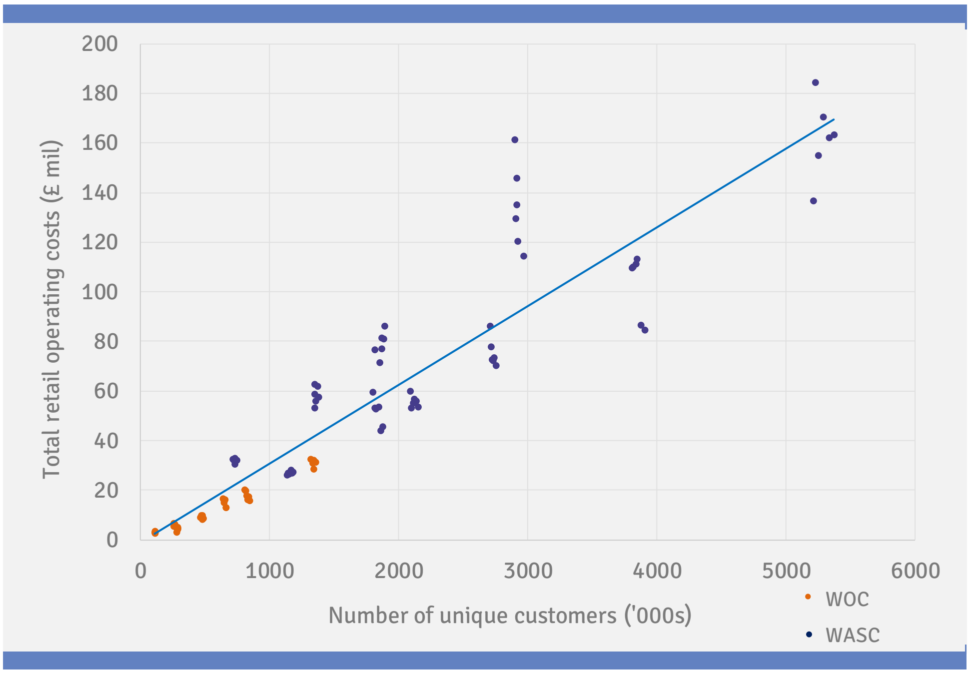
In the following, we set out a *summary* of our descriptive statistics analysis, where we outline the key cost drivers considered within the scope of our work. A more comprehensive descriptive statistics analysis, in which we described full details of the various measures examined for each driver, was provided to Bristol and Wessex separately. In reviewing the following, it is important not to 'over-interpret" individual scatterplots or correlations and use them prescriptively to arrive at generalised econometric models. Rather, our approach has been to review the raw data so that we can test hypotheses to *inform* our modelling approach, alongside engineering intuition.

3.3.1 Customer numbers – scale and scope

It is widely acknowledged that scale and scope related factors (i.e. customer numbers) are key cost drivers of retail activities in the water industry. With regards to scale, most obviously, companies incur higher costs in absolute terms the more customers they serve. Therefore, we expect strong, positive, relationships between total retail operating costs (but also our disaggregated measures of: bad debt related retail costs; and non-bad debt related retail costs) and customer numbers.

To illustrate this, the following figure shows the scatterplot between total retail operating costs against the number of unique customers. As is to be expected, there is an extremely close relationship between customer numbers and the absolute levels of retail costs incurred by companies.

Figure 3: Scatterplot of **total retail operating costs** against the **number of unique customers**



Source: Economic Insight

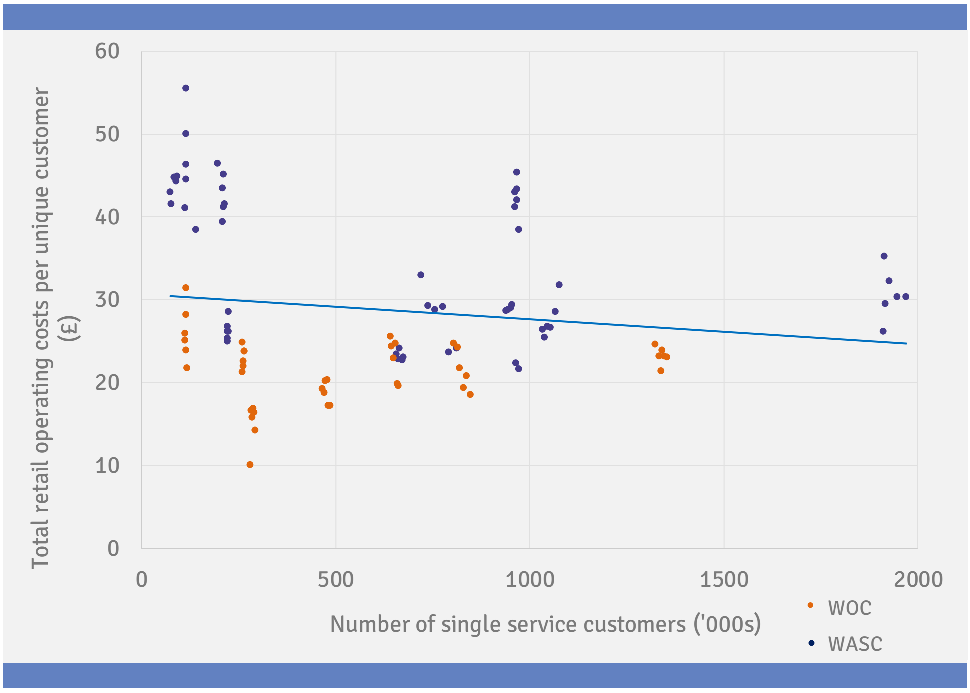
In addition to customer numbers driving the total amount of cost, there may also be economies of scope and economies of scale.

- **Economies of scale** arise because there are fixed costs inherent in providing retail services. To the extent that this is the case, one would expect to see a *negative* relationship between *unit* retail costs and customer numbers.

- **Economies of scope** arise because the costs of providing retail services to both water and wastewater customers may be less than the combined cost of providing them individually. Again, if this were the case, we would expect to see a *negative* relationship between *unit* retail costs and measures of 'scope' (i.e. the number of dual serve customers).

In relation to the former, the following figure shows a scatterplot between retail operating costs per unique customer against the number of single service customers. This shows a negative relationship, as expected.

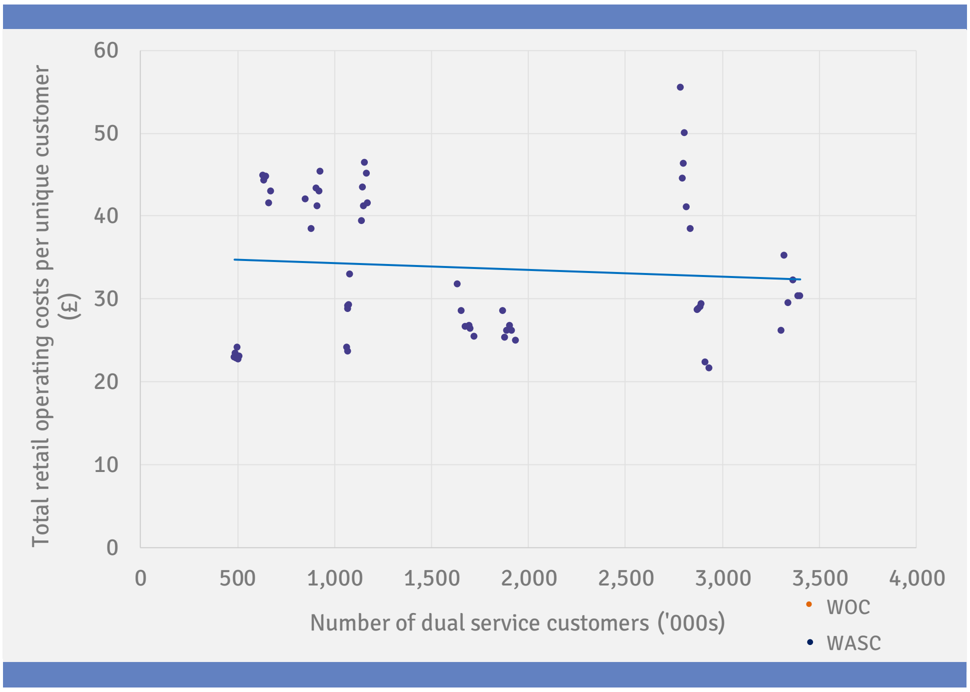
Figure 4: Scatterplot of **total retail operating cost per unique customer** against the **number of single service customers**



Source: Economic Insight

The next figure shows a scatterplot of retail costs per dual serve customer against the number of dual service customers, for WaSCs only. Also, as expected, this shows a negative relationship.

Figure 5: Scatterplot of **total retail operating cost per dual service customer** against the **number of dual service customers** (WaSCs only)



Source: Economic Insight

As shown in the table below, the correlation coefficients are *consistent with* scale and scope economies persisting for both bad debt related retail costs; and non-bad debt related retail costs.[9]

Table 5: Correlation between costs and customer number measures

Measures		Unique customers ('000s)	Single service customers ('000s)	Dual service customers ('000s)	Population ('000s)
Retail operating costs	Total cost	0.939	0.481	0.889	0.895
Retail operating costs	Unit cost	0.298	**-0.157**	**-0.090**	0.310
Bad debt related retail costs	Total cost	0.870	0.399	0.782	0.820
Bad debt related retail costs	Unit cost	0.386	**-0.097**	**-0.095**	0.371
Non-bad debt related retail costs	Total cost	0.952	0.534	0.899	0.916
Non-bad debt related retail costs	Unit cost	0.039	**-0.205**	**-0.058**	0.093

Source: Economic Insight

[9] *Note, 'consistent with' here means specifically this. One cannot directly infer that there are definitely economies of scale or scope based on these correlations alone.*

3.3.1.1 Key implications

THE DATA IS CONSISTENT WITH THE PRESENCE OF BOTH ECONOMIES OF SCALE AND SCOPE.

The importance of scale as a retail cost driver is uncontentious – and, indeed, is consistent with Ofwat's PR14 approach (i.e. by setting retail costs on a weighted average unit basis, the regulator explicitly allowed for scale). Similarly, the fact that the data is *consistent with* economies of scope is also expected – and again is consistent with prior regulatory approaches (at PR14 Ofwat applied an adjustment factor to average retail unit costs to reflect this).

3.3.2 Meter penetration

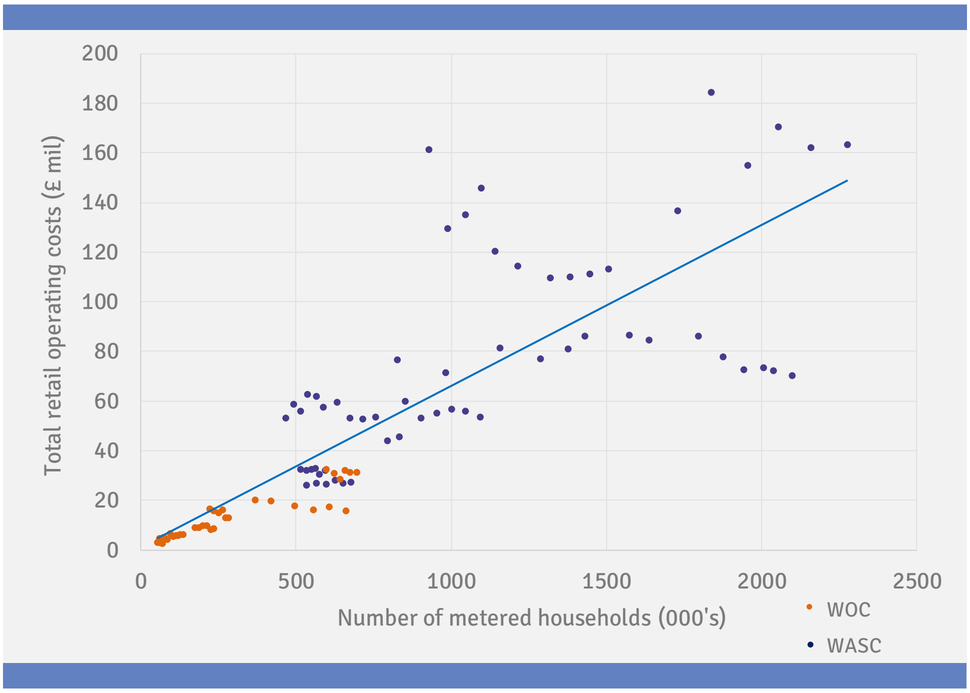
Meter penetration refers to the extent to which a company serves metered, versus unmetered, households. Clearly, in absolute terms, it is more expensive for companies to serve metered households, relative to unmetered households, because the former requires companies to undertake additional activities (i.e. to physically read meters).

Consistent with the following scatterplots, one would therefore expect there to be a positive relationship:

- between total retail operating costs and the number of metered households; and
- between retail costs *per customer* and the *proportion* of metered households.

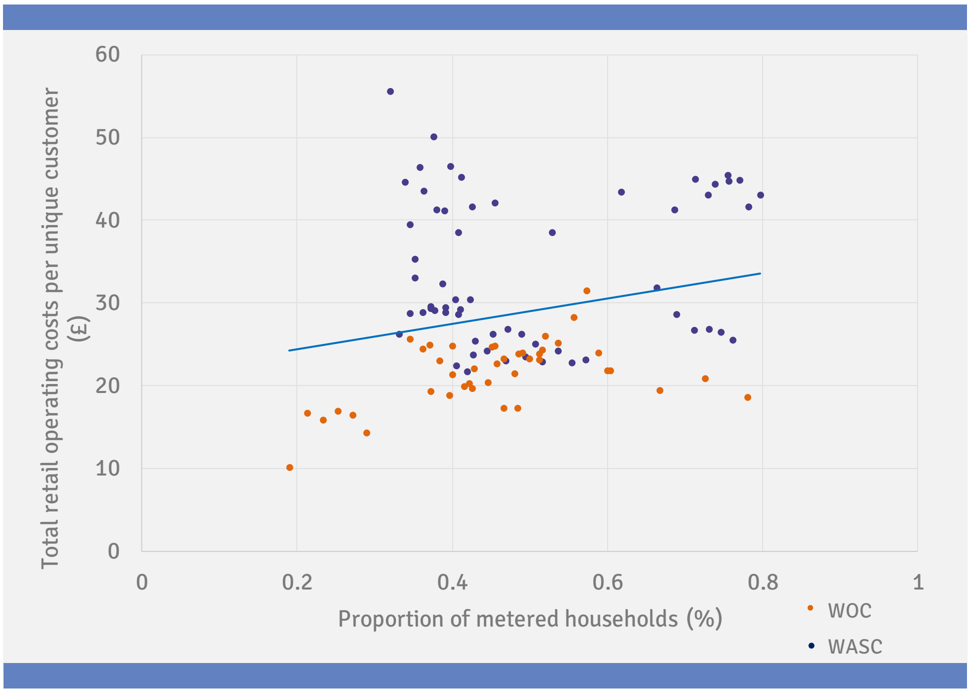
METER PENETRATION IS FOUND TO BE A CLEAR DRIVER OF RETAIL COSTS. FOR MODELLING PURPOSES, WE ADVOCATE USING THE 'PROPORTION', RATHER THAN THE NUMBER, OF METERED HOUSEHOLDS, AS THE ISSUE OF 'SCALE' SHOULD BE CONTROLLED FOR SEPARATELY.

Figure 6: Scatterplot of **total retail operating costs** against the **number of metered households**



Source: Economic Insight

Figure 7: Scatterplot of **total retail operating costs per unique customer** against the **proportion of metered households**



Source: Economic Insight

The following table summarises the correlation coefficients relating to our measures of meter penetration and retail costs – which accord with our prior expectations.

Table 6: Correlation between meter penetration measures and retail operating costs

Measures		Metered households (000's)	Metered households (%)
Total operating cost	Total cost	0.839	**-0.128**
Total operating cost	Unit cost	0.305	0.231
Bad debt	Total cost	0.783	**-0.105**
Bad debt	Unit cost	0.409	0.202
Non-bad debt related retail costs	Total cost	0.845	**-0.144**
Non-bad debt related retail costs	Unit cost	0.017	0.199
Meter reading cost	Total cost	0.851	**-0.041**
Meter reading cost	Unit cost	0.295	0.405

Source: Economic Insight

3.3.2.1 Key implications

The above descriptive statistics are, as expected, consistent with meter penetration being a driver of retail costs. We note this also accords with Ofwat's PR14 approach to household retail, where the regulator set separate average costs allowances for 'metered' and 'unmetered' customers. Because the issue of 'scale' (i.e. customer numbers) should be captured separately within the modelling approach (as described above) we consider it most appropriate to use the 'proportion of metered households' as an explanatory variable within a generalised modelling framework. As set out previously, however, as with most meter cost related drivers, because metering only accounts for a relatively small proportion of the cost stack, meter penetration may, or not, be significant when included within broader econometric models.

3.3.3 Population or meter density

Population / meter density refers to how closely located a company's customers (or meters) are to one another, on average. Clearly, this could impact retail costs – most obviously, the meter reading element of costs – because:

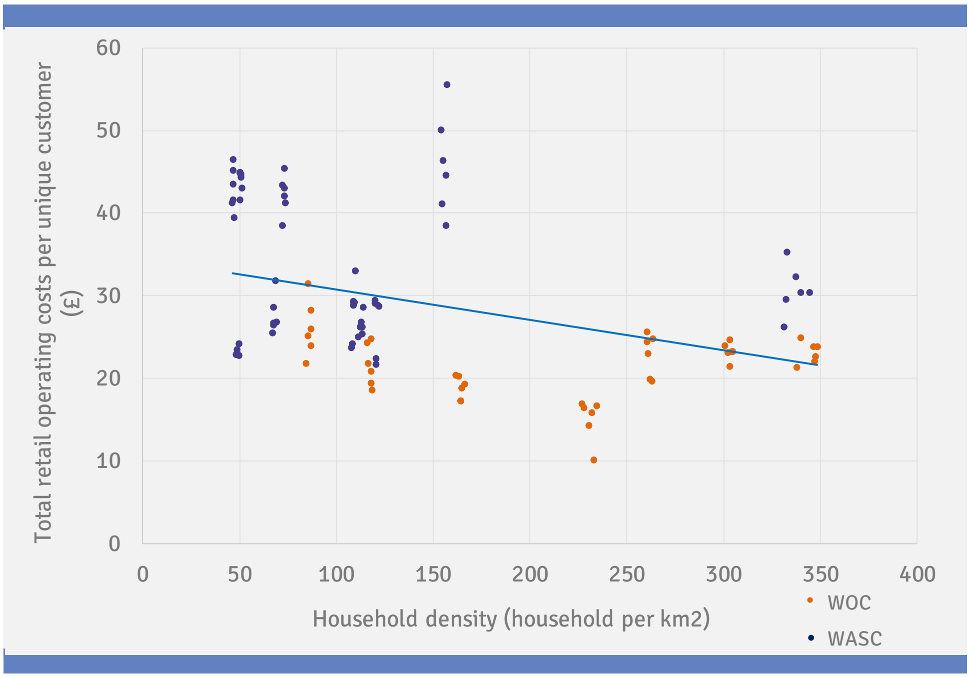
- greater distances between customers / meters will make it more time-consuming to read meters, resulting in higher labour and fuel costs; and
- relatedly, greater distances may result in increased wear on tear on meter reading equipment.

Accordingly, we expect to see generally negative relationships between density measures and unit (i.e. per customer) measures of retail costs. Importantly, because 'density' is a function of the geographic configuration of company supply zones, it can generally be considered to be outside of efficient management control.

In the following, we highlight selected analyses to illustrate the key points. The following scatterplots show:

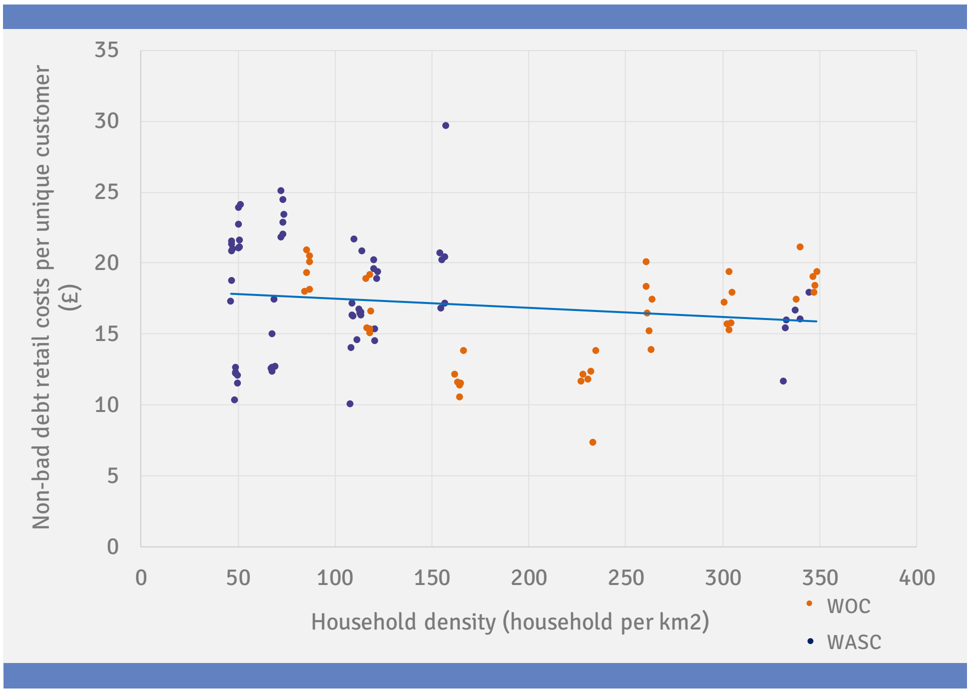
- **Retail operating costs per unique customer**, against density (households per square km).
- **Non-bad debt related retail costs per unique customer,** against density (households per square km).

Figure 8: Scatterplot of **retail operating costs per unique customer** against the ratio of: **household per sq km**



Source: Economic Insight

Figure 9: Scatterplot of **non-bad debt costs per unique customer** against the ratio of: **households per sq km**



Source: Economic Insight

Both of the above figures are consistent with unit retail costs declining with 'density' measures. As summarised in the following table (setting out the related correlation coefficients) we found that this holds across a range of alternative measurement approaches.

Table 7: Correlation between retail operating costs and population density measures

Measures		Population density (people per km²)	Household density (household per km²)	Metered households to supply area (km²)
Total operating cost	Total cost	0.070	0.037	0.364
Total operating cost	Unit cost	**-0.318**	**-0.389**	**-0.193**
Bad debt	Total cost	0.048	0.019	0.338
Bad debt	Unit cost	**-0.386**	**-0.438**	**-0.185**
Non-bad debt related retail costs	Total cost	0.086	0.053	0.368
Non-bad debt related retail costs	Unit cost	**-0.085**	**-0.164**	**-0.136**
Meter reading cost	Total cost	0.156	0.088	0.470
Meter reading cost	Unit cost	**-0.075**	**-0.196**	0.113

Source: Economic Insight

Table 8: Correlation between retail operating costs and population density measures

Measures		Population relative to mains length	Households relative to mains length	Metered households per mains length km
Total operating cost	Total cost	0.076	0.129	0.350
Total operating cost	Unit cost	**-0.302**	**-0.283**	0.188
Bad debt	Total cost	0.039	0.090	0.335
Bad debt	Unit cost	**-0.372**	**-0.352**	0.206
Non-bad debt related retail costs	Total cost	0.108	0.159	0.344
Non-bad debt related retail costs	Unit cost	**-0.073**	**-0.063**	0.091
Meter reading cost	Total cost	0.167	0.205	0.500
Meter reading cost	Unit cost	**-0.034**	**-0.043**	0.495

Source: Economic Insight

3.3.3.1 Key implications

Across all the density measures, we see that there is a negative relationship between unit retail cost measures and density, which accords prior expectations and is intuitively sensible. This, then, is consistent with including some form of density measure within a generalised modelling framework. We suggest that density measures in terms of square km of supply area and length of mains are, in principle, valid measures to consider. However, of the two, the length of main approach may more accurately reflect the way in which density drives metering related retail costs.

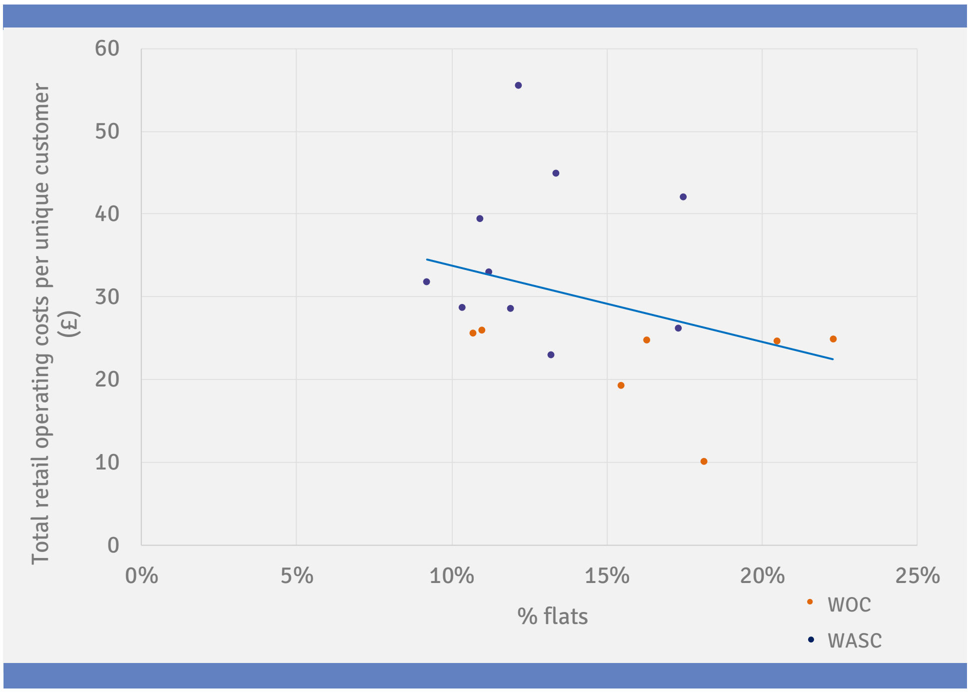
3.3.4 Housing stock

By 'housing stock', we refer to the configuration or mix of 'housing types' within and across company supply areas. In principle, there are several reasons as to why this might impact retail costs in the water industry – mainly in relation to metering. For example, it might be more time-consuming and / or 'difficult' to read meters in relation to detached houses, relative to, say, flats. For example, we understand that, in blocks of flats, meters for multiple properties can be co-located, reducing read time. If that were the case, one might find that staff (and potentially other) related costs will be lower 'per meter read' for flats, relative to other housing types.

However, as it is not unambiguously clear how housing stock might impact costs, we have no strong a priori views on this matter – and so we explored a number of measures. For summary purposes, however, the next figure shows a scatterplot of total retail operating costs per unique customer against the percentage of flats in company supply areas.

Importantly, the configuration of housing stock is outside of management control – and so, if it were considered to be a cost driver – there would be an 'in principle' rationale for capturing it within cost assessment.

Figure 10: Scatterplot of **retail operating costs per unique customer** against the **percentage of flats**



Source: Economic Insight

The following table contains details of the correlation coefficients for the various combinations of housing stock and cost measures.

Table 9: Correlation between costs and customer number measures

Measures		Social housing (%)	Flats (%)
Total operating cost	Total cost	-0.006	-0.268
Total operating cost	Unit cost	-0.191	-0.334
Bad debt	Total cost	-0.002	-0.201
Bad debt	Unit cost	-0.202	-0.340
Non-bad debt related retail costs	Total cost	-0.011	-0.316
Non-bad debt related retail costs	Unit cost	-0.104	-0.238
Meter reading cost	Total cost	-0.019	-0.180
Meter reading cost	Unit cost	0.003	-0.108

Source: Economic Insight

THE EVIDENCE IS CONSISTENT WITH INCLUDING A MEASURE OF HOUSING STOCK WITHIN A GENERALISED MODELLING FRAMEWORK. HOWEVER, GIVEN THE POTENTIAL INTERACTIONS WITH OTHER COST DRIVERS, IT MAY OR MAY NOT PROVE TO BE STATISTICALLY VALID.

3.3.4.1 Key implications

There does appear to be a relationship between both total and unit retail costs and housing stock. Most notably, the percentage of flats in company supply areas is negatively correlated with all retail cost measures – which accords with our priors. It is important to recall that the correlation between housing stock and costs, viewed in isolation as above, may be impacted by other retail (particularly metering) related cost drivers. For example, if the percentage of flats is correlated with density, then – in part – the negative coefficient may also reflect the 'depressing' impact density has on costs. Conversely, if the percentage of flats is correlated with congestion (see below) which has an 'increasing' impact on costs, then this might be mitigating the apparent impact of housing stock reported above. Consequently, it is difficult to say, ex-ante, whether one should expect this measure to be statistically valid within an econometric model setting. Nonetheless, we consider that the evidence here is supportive of including a housing stock measure within a generalised modelling framework (particularly as it is self-evidently outside of efficiency company control). We further find that the % of flats within company supply areas is the more credible metric.

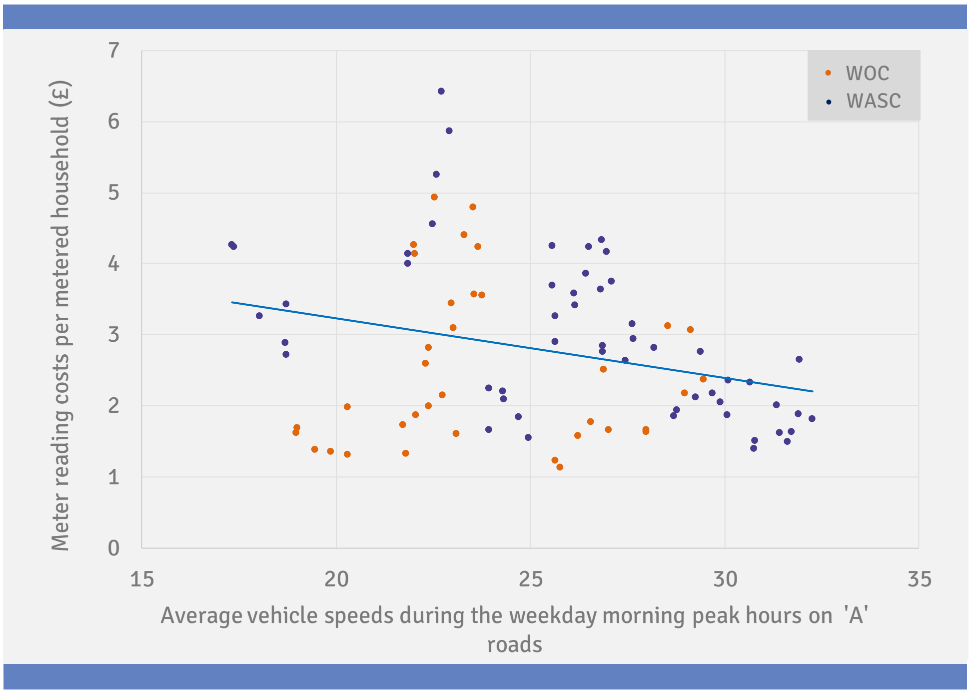
3.3.5 Congestion

LOWER AVERAGE TRAFFIC SPEEDS ARE ASSOCAITED WITH HIGHER UNIT COSTS.

In principle, road or traffic congestion could impact retail costs. Most obviously, the more congested roads are, the higher meter reading costs are likely to be (because more time is required 'per read' – thus impacting labour costs, but also potentially, fuel costs). Other retail activities that include a labour element that requires travel within company supply areas could also be impacted. Relatedly, for the purpose of cost assessment, underlying congestion can generally be expected to be outside of efficient management control. Nonetheless, companies can clearly mitigate the impact of congestion in how they optimise their metering reading processes – and so we would expect Ofwat to pay attention to this.

We have explored several ways of measuring congestion. These include vehicle flow rates and various alternative measures of average speeds. For summary purposes, the following figure shows the scatterplot of 'metering costs per metered customer' against average traffic speed (measured as morning peak miles per hour). As expected, this shows lower average speeds are associated with higher unit retail costs.

Figure 11: Scatterplot of **metering costs per metered customer** against the **average traffic speed (morning peak, MPH)**



Source: Economic Insight

The above measure of average speed was the same one used by Bristol in its congestion special factor cost claim at PR14, which was accepted by Ofwat.[10]

The following table summarises the relevant correlation coefficients for the various congestion / cost combinations we examined.

Table 10: Correlation between retail operating costs and congestion measures

Measures		Congestion (average daily flow)	Average speed (mph)	Average speed A road weekday morning peak
Total operating cost	Total cost	**-0.112**	**-0.310**	**-0.241**
Total operating cost	Unit cost	**-0.485**	0.099	0.050
Bad debt	Total cost	**-0.152**	**-0.266**	**-0.243**
Bad debt	Unit cost	**-0.560**	0.162	0.123
Non-bad debt related retail costs	Total cost	**-0.067**	**-0.335**	**-0.225**
Non-bad debt related retail costs	Unit cost	**-0.182**	**-0.044**	**-0.083**
Meter reading cost	Total cost	0.036	**-0.325**	**-0.240**
Meter reading cost	Unit cost	**-0.066**	**-0.001**	0.036
Meter reading cost	Cost per metered household	**-0.001**	**-0.231**	**-0.280**

Source: Economic Insight

3.3.5.1 Why time of day matters

IT IS IMPORTANT TO FACTOR 'TIME OF DAY' INTO ANY CONGESTION MEASURE, TO AVOID INAPPROPRIATELY BENEFITTING INEFFICIENT COMPANIES.

When considering congestion as a potential cost driver, it is important to take 'time of day' into account. Specifically: suppose there were two companies, both with the same overall average speed – but that one had materially lower average speed during working hours (i.e. when meter reading takes place) than the other. In such circumstances, if one only included the overall average speed in a cost model, both companies would be treated equally – even though one genuinely faces greater congestion of relevance to meter reading – and therefore, higher costs. This would clearly be inappropriate.

Following from the above, it should be non-contentious that 'time of day' should be factored into any congestion measure. Further, ideally one would wish to use a measure of average speed *during working hours by locality.* In practice, we are not aware that such a measure exists. Consequently, one can think of the 'morning peak' measure (which is available on a localised level) as being a proxy for this. Helpfully, empirical evidence suggests that this is, in fact the case – and moreover, Ofwat accepted this at PR14 in relation to Bristol Water's congestion related special factor cost claim. Importantly, if one is satisfied that the 'morning peak' measure is a reasonable proxy for working hours average speeds, then there should be no material concerns regarding the measure not reflecting companies' ability to optimise meter reading to avoid the morning peak.

[10] *See 'Bristol Water Representation on the PR14 Draft Determination: Appendices.' (2014).*

3.3.5.2 Key implications

We find that, as expected, if appropriately measured, congestion is positively associated with retail costs (or, in the case of 'average speed', is negatively correlated). Of the measures assessed, we consider measures of 'average speed' to be preferable to flow or other congestion metrics. This is because it avoids conflating scale with the underlying issue of interest here. More specifically, and as summarised above, we consider the peak morning speed measure to be most suitable (as it acts as a reasonable proxy for congestion in working hours).

Given the above, we support including a congestion variable within a generalised modelling framework. As previously discussed, because this variable likely interacts with the issues of 'density' and 'housing stock' it is not clear, ex-ante, whether it is likely to be statistically valid within any specific econometric model.

3.3.6 Regional wages

'When one considers the specific labour-related activities associated with the retail element of the water value chain, there is – in our view – no strong ex-ante reason to suppose that regional wage variation should be controlled for.'

In the context of retail cost assessment, the issue of 'regional wages' refers to the fact that costs might vary across companies due to underlying differences in regional labour market conditions, which are outside of efficient management control. While, in principle, one might suppose that such effects *may* arise, when one considers the specific labour-related activities associated with the retail element of the water value chain, there is – in our view – no strong ex-ante reason to suppose that regional wage variation should be controlled for.

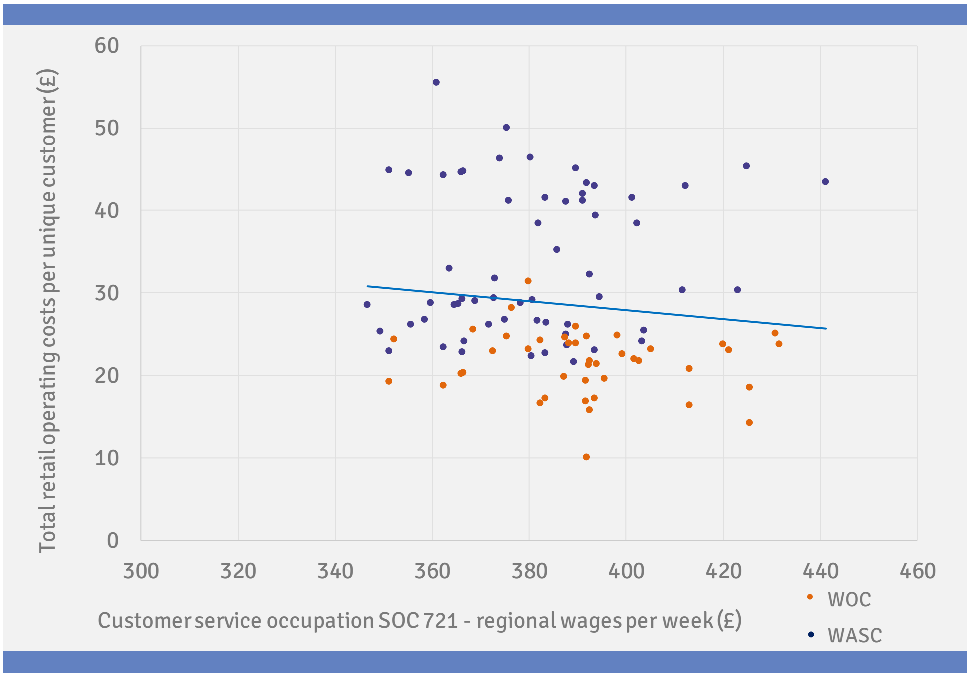
The primary reason for this is that many labour-intensive retail activities can be outsourced – not only within the UK – but internationally. Consequently, while it is true that labour costs do vary regionally – this is not particularly relevant to the efficient retail costs a water company would incur. This is consistent with Ofwat's position, as referenced previously in our report. Specifically, in its draft PR19 methodology, Ofwat states: *"With the exception of metering, which is around 5% of total retail costs, we consider that the impact of regional labour costs can be substantially mitigated or removed in retail activities."*[11]

Nevertheless, to inform our modelling approach, we examined a range of regional wage metrics to evaluate their relationship with measures of retail costs. The following figures provide what we consider to be the most pertinent analysis – namely scatterplots showing the relationship between average weekly wages for occupations relevant to retail (customer service) and: (i) total retail operating costs per customer; and (ii) metering costs per customer.

---

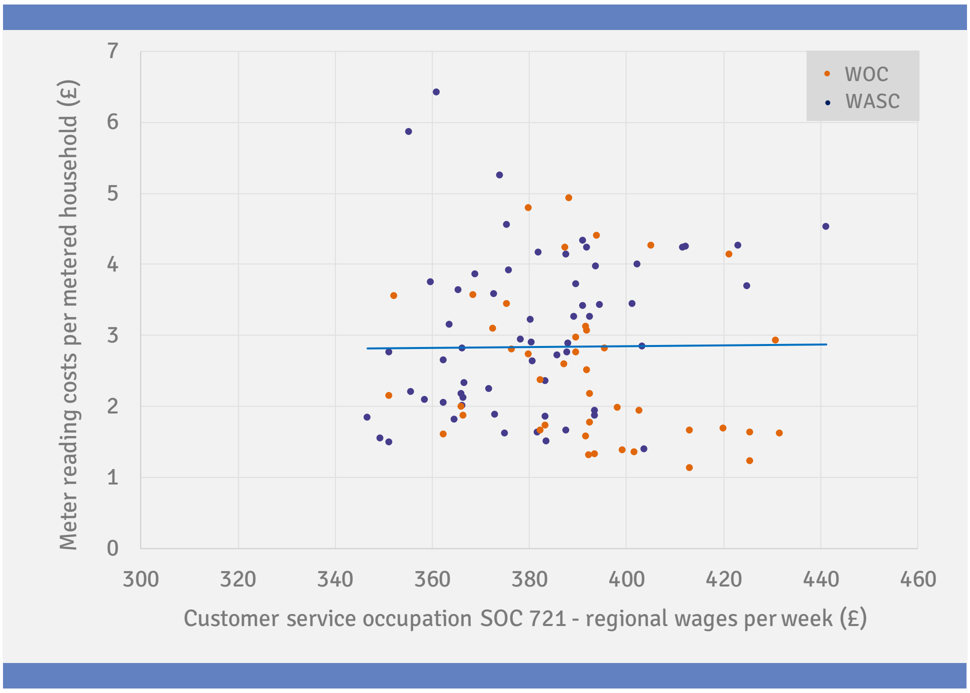
[11] *'Delivering Water 2020: Consulting on our methodology for the 2019 price review.' Ofwat (2017); page 184.*

Figure 12: Scatterplot of **total retail operating costs per unique customer** against **average weekly regional wages for customer service occupations**



Source: Economic Insight

Figure 13: Scatterplot of **metering costs per metered customer** against **average weekly regional wages for customer service occupations**



Source: Economic Insight

Relating to the above, we find that:

- Total retail operating costs per customer are in fact *negatively* related to the relevant regional wage measure. This is consistent with any regional wage effect being sufficiently small that it is not apparent in an analysis of the overall retail cost stack.

- When we look specifically at metering costs, there is a *slight positive correlation.* However, this is sufficiently small such that it is unlikely that there is any real effect, nor one that would be valid within an econometric model.

The above results accord with priors, and Ofwat's current position: namely, that in relation to retail activities, any regional wage effect is likely to be immaterial. The following table contains a summary of the relevant correlation coefficients.

Table 11: Correlation between costs and measures of regional wages

Measures		Overall average weekly wages (regional) (£)	Customer service occupation weekly (regional) SOC 721 (£)	Sales-related occupations weekly wage (regional) SOC 712 (£)
Total operating cost	Total cost	**-0.033**	**-0.075**	0.010
Total operating cost	Unit cost	**-0.279**	**-0.115**	**-0.200**
Bad debt	Total cost	**-0.033**	**-0.046**	0.029
Bad debt	Unit cost	**-0.309**	**-0.144**	**-0.162**
Non-bad debt related retail costs	Total cost	**-0.030**	**-0.099**	**-0.010**
Non-bad debt related retail costs	Unit cost	**-0.127**	**-0.024**	**-0.193**
Customer service cost	Total cost	0.025	**-0.051**	0.071
Customer service cost	Unit cost	0.063	0.034	0.066
Meter reading cost	Total cost	0.108	0.048	0.061
Meter reading cost	Unit cost	0.020	0.127	**-0.094**

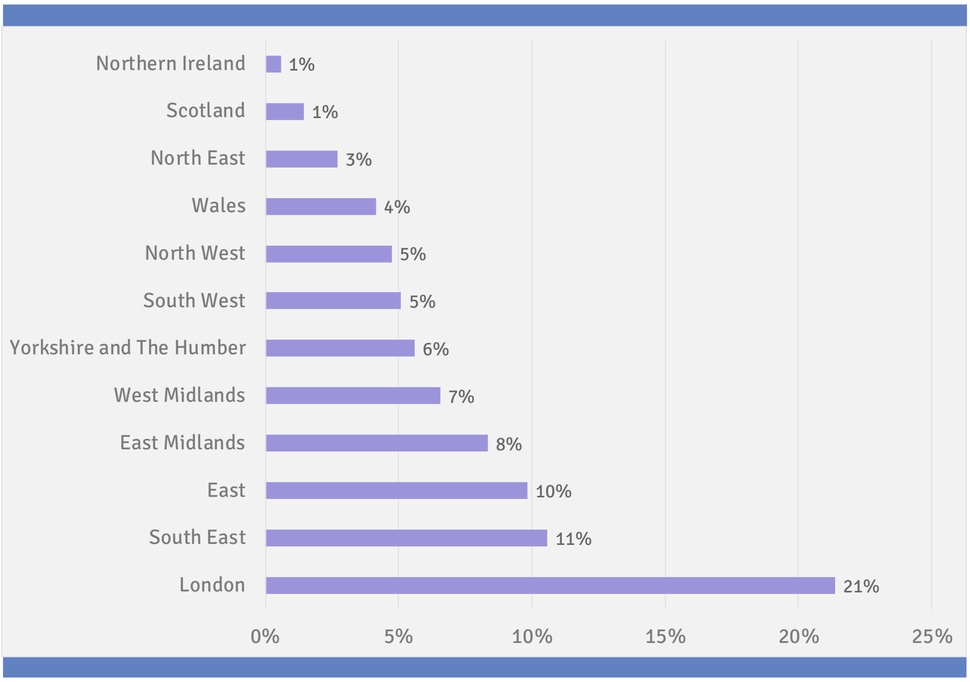
Source: Economic Insight

The above analysis ultimately goes to the question of whether the labour markets for water retail activities are 'regional' in scope, or are broader. In the round, the evidence is consistent with them being broader – and therefore, the data does not show any strong regional wage effect.

The main explanation for this is likely to be that, as indicated previously, many labour-intensive retail activities can be 'outsourced' from a geographic perspective. For example, customer service agents do not have to be located within a company's supply area.

In addition, even to the extent that there are certain roles that require workers to undertake activities within supply areas, this does not mean that companies must necessarily employ workers within those supply areas. This is because there is a difference between *place of work,* and *place of residence.* Put simply, some 'regions' may be able to draw on a wider supply base of workers that extends out beyond regional boundaries. This is illustrated in the following chart, which shows the percentage of workers, by region, who normally reside outside of their region of work.

Figure 14: Percentage of workers normally resident outside of their region of work



Source: NOMIS, ONS, Census – relates to 2011

It is important to understand that the above chart does not provide further evidence of whether there is, or is not, a regional wage effect for water retail (as, by definition, wages will reflect the balance of demand and supply, which will of course reflect commuting propensity). Rather, it should be seen, alongside outsourcing, as part of the explanation as to *why* no strong regional effect is found in the data.

3.3.6.1 Key implications

The descriptive statistics analysis is generally not consistent with regional wage measures being an important driver of retail related costs for the water industry. This largely accords with priors. Namely, that the scope of 'relevant' labour markets for retail activities is, in most cases, likely to be broader than regional (most obviously because of the ability of companies to 'outsource' certain activities, but also, potentially, because firms may be able to draw on a pool of workers that is 'wider' than their supply area).

Given the above, our expectation is that regional wage variation should not be included within a cost assessment model for household retail at PR19. For completeness, however (and recognising the 'in principle' connection to metering costs) we subsequently test what we consider to be the most relevant regional wage measure within our generalised modelling framework. We subsequently find this to be insignificant.

3.3.7 Regional socioeconomic performance (deprivation)

Regional socioeconomic performance (or deprivation) is likely to impact the bad debt element of household retail costs, because in areas with poorer socioeconomic indicators, customers are more likely to fall into arrears, or default. It might be that:

- more deprived customers are more exposed to financial shocks, and are more likely to enter financial distress; and / or
- additional issues associated with deprivation as a 'wider concept' – say relating to financial literacy, may also impact the propensity of customers to fall into arrears etc.

Clearly, geographic variance in socioeconomic performance is outside of efficient management control and so, at PR14, the *principle* of allowing for this was accepted as part of the special cost factor claim process[12] (in practice, claims were subject to Ofwat's evidential hurdles, including demonstrating that companies were making best endeavours to minimise the impact of deprivation on their debt related costs).

At PR14, extensive analysis was undertaken of the potential impact of deprivation on bad debt costs – with various models proposed and examined by stakeholders. Two prominent issues that were debated at the time included: (i) what is the most appropriate way of measuring deprivation *for the purpose of identifying its impact on company costs;* and (ii) what is the appropriate way of statistically modelling *projected* allowed costs over the price control? For our purposes, only the first of these two issues is relevant.

Several measures of regional socioeconomic performance / deprivation could be used to inform cost assessment in practice. One of the main measures, and the one primarily used at PR14, is the Index of Multiple Deprivation. The IMD (in England) is published by the Department for Communities and Local Government and is the *"official measure of relative deprivation for small areas (or neighbourhoods) in England."*[13]

The main benefits of using the IMD are that:

- It represents the UK's 'official' measure of deprivation.
- It encapsulates a 'broad' concept of deprivation, in that it provides an 'overall score' for deprivation, of which sub-components include: income; employment; education; health; crime; barriers to housing; and living environment.
- It is available on a localised (super output area) level – which therefore provides extensive variation that can be used in statistical analysis to better inform the extent to which it impacts debt costs.

In practice, the first and second points are perhaps more important to efficiency benchmarking (as generally in a panel dataset, there is likely to be sufficient variation to identify relationship across company supply areas, without necessarily the need to utilise lower level data).

[12] *For example, see 'Final price control determination notice: company-specific appendix – South West Water.' Annex 2. Ofwat (2014).*
[13] *'The English Index of Multiple Deprivation (IMD) 2015 – Guidance.' Department for Communities and Local Government (2015).*

The downside of the IMD, however, is that (in relation to the *'overall IMD score'*) it is not available on a consistent basis across England and Wales. This means that utilising IMD data for efficiency benchmarking might require:

- One to omit Welsh Water as a comparator, thus reducing the number of observations (and potentially resulting in efficient costs being over or understated for the industry).
- Creating an estimated version of the measure for Wales (which might be especially important if one takes the view that the various dimensions of deprivation have distinct impacts on debt-related costs that need to be captured).
- Using a sub measure of the IMD – most obviously, IMD income. The advantage of this approach is that this metric is constructed similarly for England and Wales – measuring the proportion of population entitled to the same benefits.[14] In addition, as IMD income is likely to be a key driver of (and will be highly correlated with) overall deprivation, this would seem to be a valid and credible approach. One potential issue, however, is that (as above) it may not reflect the *broader dimensions* of deprivation – and how they impact debt-related costs. The implications of this for cost assessment would vary, depending on whether one thought any broader dimensions of deprivation affected debt-related costs: (i) *across the industry;* or (ii) only in relation to *specific companies.*

WHILST THE IMD IS NOT AVAILABLE AS A TIME SERIES, IT CAN BE COMPLEMENTED WITH VARIOUS OTHER PUBLICLY AVAILABLE SOCIOECONOMIC PERFORMANCE MEASURES, WHICH ARE AVAILABLE OVER TIME.

Another issue with the IMD data is that it is not recorded on a consistent basis over time. As clearly bad debt costs do vary over time (as well as across companies) as a matter of principle, it is arguably preferable to use socioeconomic measures for which a consistent time series is available. In practice, whether this matters depends on the extent to which the *differences* in socioeconomic performance across companies in fact vary over the time periods in question, or are relatively stable. If the latter is the case, then in modelling the distinction between time-varying and static measures of socioeconomic performance will be less important. In addition, to the extent that differences in socioeconomic performance across companies do vary over time, in practice the IMD measures could be supplemented with other measures, which do vary.

Related to the above, in addition to the IMD, there are numerous other measures of socioeconomic performance at a regional level. Consequently, in our descriptive statistics analysis, we have reviewed and considered a wide range of potential metrics. These are summarised in the following table.

[14] *We have carefully reviewed the Technical Reports describing the construction of IMD income in England and Wales respectively. We note that the same benefit types (i.e. the indicators used to measure the proportion of population meeting the 'low income criteria') are listed in both. In relation to one specific indicator (child tax credits) the measure captures the count of families with an equivalised income below 60% of the median. As median incomes will vary between England and Wales, this will, therefore, not be comparable like-for-like. However, we nonetheless consider that IMD income can be compared across England and Wales because: (i) this issue only applies to one specific indicator <u>within</u> the IMD income score; and (ii) the Technical Reports further state that even this one indicator only includes population <u>not already captured by other criteria</u>. As such, the impact on the IMD income score of median incomes differing between England and Wales is likely to be negligible (although one would expect IMD income scores to be somewhat understated for Wales). We further note that previous academic studies have utilised IMD income across England and Wales, consistent with our view that they are comparable.*

Table 12: Potential measures of regional socioeconomic performance

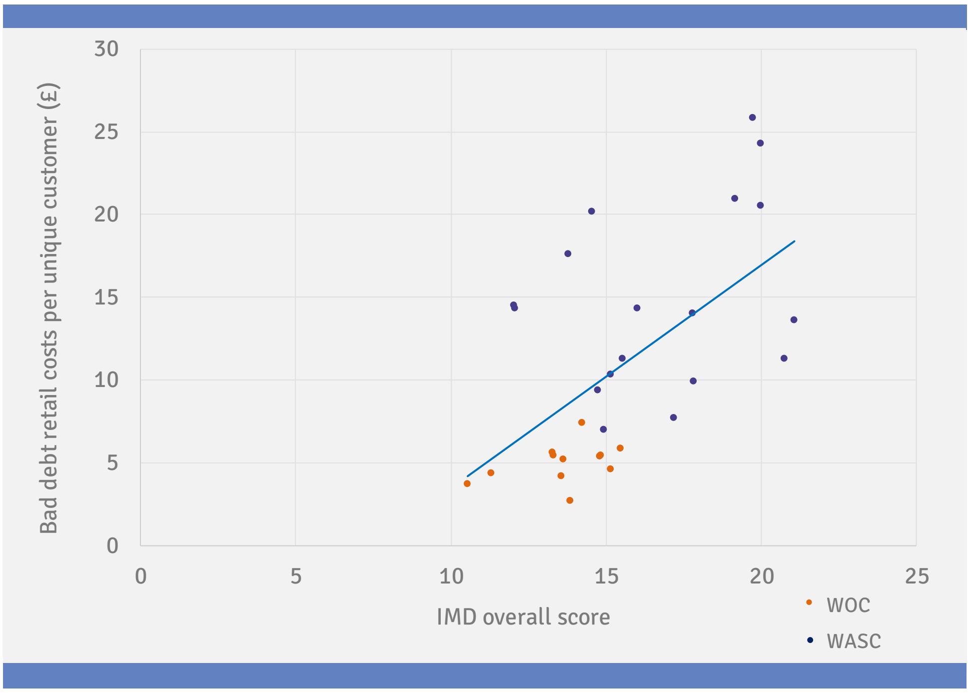
Variable	Definition	Source
JSA	Number of people receiving Job seeker's allowance	ONS
JSA rate	% of people on receiving JSA	ONS
Number of workless households	Number of households where no member of the house is employed	ONS
Percentage of workless households	Proportion of total households where no member of the house is employed	ONS
Unemployment level	Number of people that are unemployed	ONS
Unemployment rate	Rate of unemployment	ONS
Regional wage	Average regional pay	ONS
House price to income ratio	Median house price by local authority district	ONS
Property repossession rate	Number of property repossessions divided by the number of households	ONS
IMD overall	Overall IMD score	ONS
IMD income	IMD income score	ONS
IMD employment	IMD employment score	ONS
IMD education	IMD education score	ONS
IMD health	IMD health score	ONS
IMD crime	IMD crime score	ONS
IMD barriers	IMD barriers score	ONS
IMD living	IMD living score	ONS

Source: Economic Insight

As we are addressing 'scale' elsewhere, it is most helpful to examine how the socioeconomic measures relate to bad debt related retail costs on a unit basis. In the following we highlight some of the key findings.

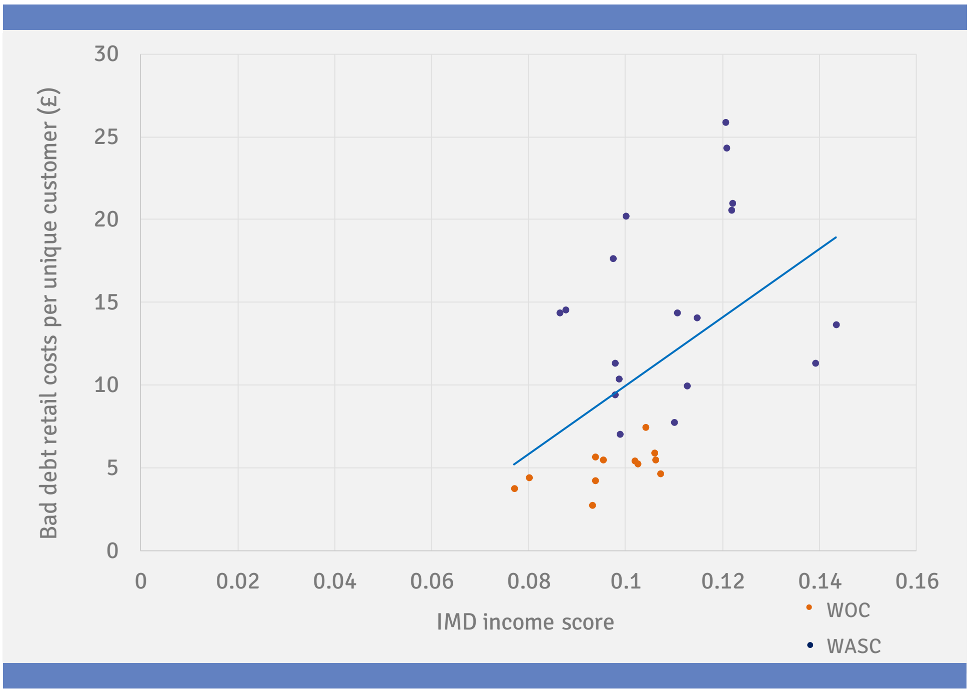
Firstly, the next two scatterplots show bad debt related retail costs per unique customer against (i) total IMD scores; and (ii) income related IMD scores.

Figure 15: Scatterplot of **bad debt related retail operating costs per unique customer** against **total IMD score**



Source: Economic Insight

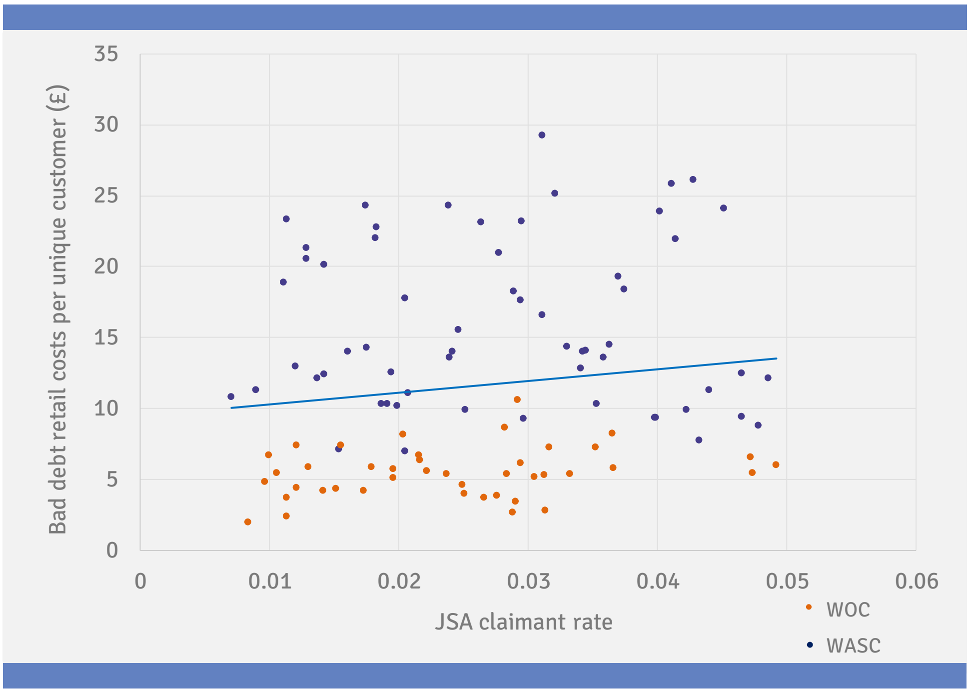
Figure 16: Scatterplot of **bad debt related retail operating costs per unique customer** against **total IMD <u>income</u> score**



Source: Economic Insight

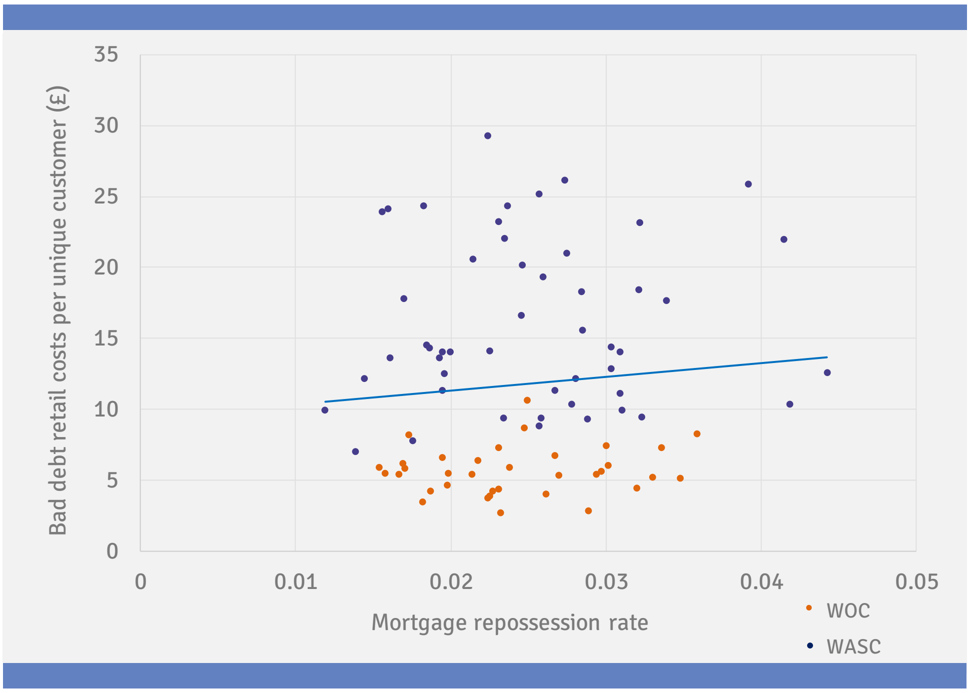
Both of the above show positive correlations indicating that, as expected, higher deprivation is associated with increased bad debt costs. The IMD income measure is shown as it allows us to include Wales. Interestingly, however, our descriptive statistics suggest that various other socioeconomic performance measures also demonstrate a relationship with bad debt costs. For example, the following figures show the same scatters, but using: (i) JSA claimant rate; (ii) property repossession rate; and (iii) house price to income ratios.

Figure 17: Scatterplot of **bad debt related retail operating costs per unique customer** against **JSA claimant rate**



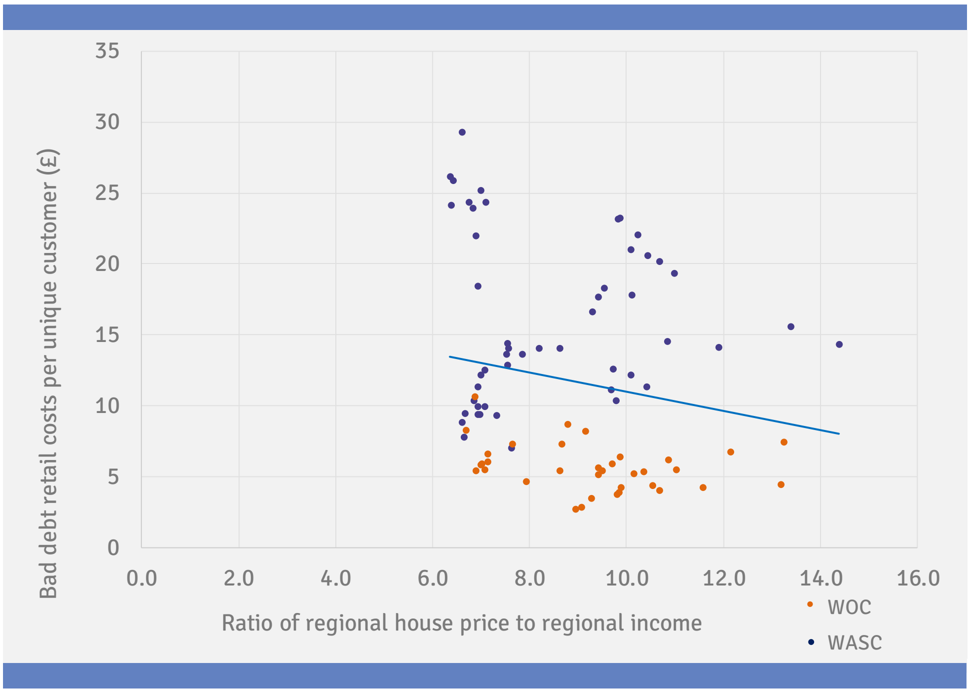
Source: Economic Insight

Figure 18: Scatterplot of **bad debt related retail operating costs per unique customer** against **property repossession rate**



Source: Economic Insight

Figure 19: Scatterplot of **bad debt related retail operating costs per unique customer** against **house price to income ratio**



Source: Economic Insight

Note, that although the correlation between the ratio of house price to income with bad debt costs per unique customer is negative, the conclusion drawn is the same as other deprivation indicators. Deprivation is inversely related to the ratio of house price to region income because, for a given income level, an individual with a higher valued house is wealthier.

The main point to highlight regarding the above, therefore, is that all of the alternative measures show the expected correlation with bad debt costs (per unique customers). Consequently, while at PR14 the focus was on IMD, it is worth noting that various metrics show a similar pattern.

Table 13: Correlation coefficients between bad debt related retail costs per unique customer and various measures of socioeconomic performance (deprivation)

Measures		JSA (%)	Unemployment (%)	Weekly average wage (£)	Workless households (%)	Property repossessions (%)	House price to income ratio
Bad debt	Total cost	0.195	0.211	-0.033	0.292	0.039	-0.154
Bad debt	Unit cost	0.137	0.080	-0.309	0.418	0.124	-0.384

Source: Economic Insight

Table 14: Correlation coefficients between bad debt related retail costs per unique customer and IMD measures

Measures		IMD overall	IMD income	IMD employment	IMD education	IMD health	IMD crime score	IMD barriers	IMD living
Bad debt	Total cost	0.202	0.133	0.136	0.158	0.273	-0.029	-0.179	0.018
Bad debt	Unit cost	0.569	0.471	0.422	0.370	0.516	-0.454	0.108	0.458

Source: Economic Insight

3.3.7.1 Considering how socioeconomic performance might interact with costs – the tails of distributions

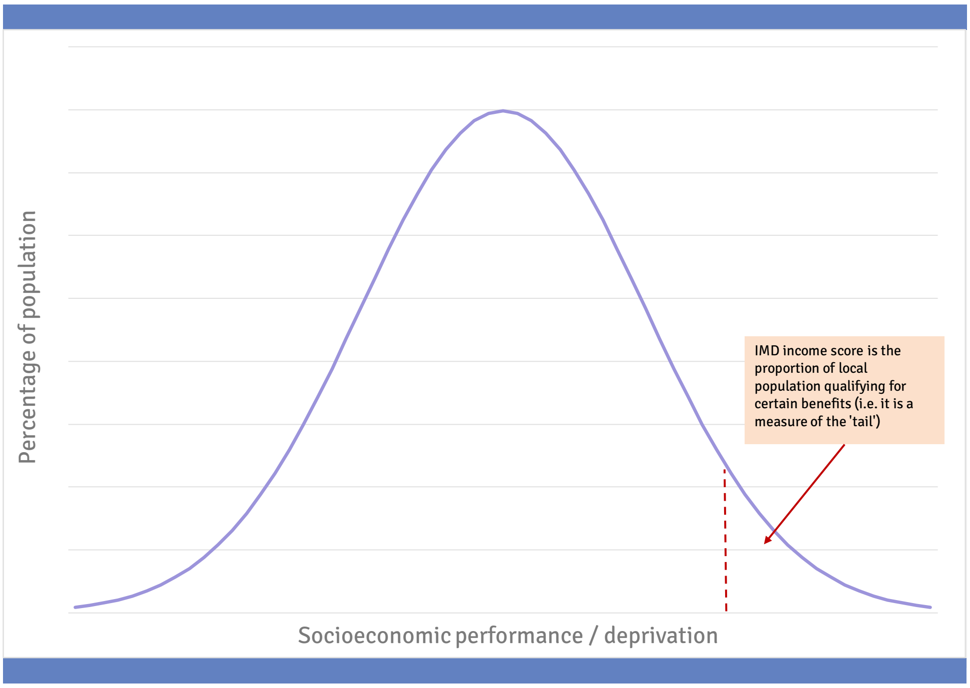
A further issue to consider in relation to incorporating socioeconomic performance / deprivation into cost assessment, is: (i) the extent to which its relationship with cost is continuous versus discontinuous; or, perhaps more pertinently, (ii) if it is continuous, whether customers who are 'most deprived' disproportionately drive cost.

- Regarding the first issue, it seems likely that the interaction of socioeconomic performance and retail costs *is continuous.* For example, in relation to debt related costs, clearly 'arrears' is a matter of degree – and so too is the 'probability of default'. Logically, therefore, customers can be thought of as being on a continuum; whereby, as their socioeconomic condition declines (i.e. deprivation gets more severe), so too their extent of arrears or probability of default increases – thus increasing retail costs in a continuous (but not, not necessarily linear) manner. Of course, above a certain level of socioeconomic performance, where customers are not in arrears at all – and have 'no' probability of default – further increases in socioeconomic performance will most likely not impact costs. Nonetheless, as a matter of principle, the interaction would seem to be continuous, up to some threshold.
- In relation to the latter issue, it is difficult to consider this independently of the measure of socioeconomic performance / deprivation being used. As a matter of principle, it may be that the key cost drivers (e.g. arrears and default probability) vary with the socioeconomic performance measure being used in a way that suggests all customers should receive equal weighing in cost assessment. On the other hand, it could also be the case that, for customers below a certain threshold on the socioeconomic performance / deprivation measure, either the extent of arrears / probability of default increases disproportionately. Intuitively, therefore, there is a rationale to explore both possibilities (i.e. including continuous variables that weight customers equally, while also examining variables that attach 'weight' to customers that are 'most deprived' – or, in other words, are in the tails of socioeconomic performance).

'If one believes that it is appropriate to include a measure that captures the proportion of people who are 'very deprived', the IMD itself already achieves this end.'

In the above context, it is important to understand the detail of how the IMD scores are calculated – particularly the IMD income measure. In summary terms, the IMD income score is a measure of the proportion of a local population that is below a deprivation threshold, which is defined based on the number of people claiming certain benefits (e.g. income support, jobseeker's allowance). However, the key point to note is that, because of the way in which it is calculated, the IMD income is itself, already a measure of the 'tail' of the deprivation – as illustrated in the following figure. Therefore, if one believes that it is appropriate to include a measure that captures the proportion of people who are 'very deprived', the IMD itself already achieves this end.

Figure 20: Illustration of how the IMD income score is a measure of 'how many' very deprived people there are within a local population



Source: Economic Insight

Following from the above, there is a separate, but related, issue of where one "draws the line" in identifying the proportion of people that are very deprived. For example, by using the IMD (either the overall IMD, or IMD income), one would, by definition, be setting the 'cut-off' based on the criteria used within the IMD. Alternatively, one could use various other socioeconomic performance / deprivation measures (such as those previously described) and define 'cut-off' points based on percentiles and so forth.

Key points to highlight regarding the above are:

- That the IMD is itself already a measure of the proportion of customers in the tail of socioeconomic performance (i.e. it captures the 'very deprived' issue).
- Consequently, alternative approaches to measuring 'extreme' deprivation are really a question of 'where' one sets the cut-off point in a distribution.
- Relating to the above, a difficulty with alternative approaches is that, the 'cut-off' point at which one observes a step change in retail costs could vary both: (i) by the measure used; and (ii) by company. A key implication of the latter is that:
  - companies may have incentives to argue for 'cut-off points' that are most in their favour from a cost assessment perspective; and
  - it would be difficult to determine whether a difference between where one observed a step change in retail costs between companies was itself due to 'inefficiency', or not.
- Following from the above, one advantage of using the IMD to define the 'cut-off' point is that it has been determined independently. Consequently, the concern regarding companies seeking to define a cut-off that most advantages them does not arise. On the other hand, a disadvantage may be that alternative cut-off points

may better reflect the *true* point at which there is a step change in industry retail costs.

3.3.7.2 Key implications

Regional socioeconomic performance (deprivation) is an intuitive driver of company bad debt costs, which is to some degree outside of efficient management control. Several potential measures all show strong positive associations with bad debt costs per customers; and so credible cases can be made to utilise various approaches. On balance, we consider that the IMD has some intrinsic value, in being the Government's *official* measure of deprivation.

A further advantage of using the IMD is that it addresses the issue of the 'most deprived' customers driving retail costs. Consequently, by including this one allows for *both* the possibility that the socioeconomic performance / deprivation relationship with cost is entirely continuous – or that retail costs may be more driven by the 'tail' of the distribution.

Given that the overall IMD measure is not available on a consistent basis across England and Wales, we consider it practical, for the purpose of this work, to focus on sub-measures of IMD, which are also available for Wales (most obviously, the IMD income score). However, recognising that deprivation is multidimensional – and that all of its aspects might affect (efficient) debt-related retail costs, we consider that from a modelling perspective, there is merit in further exploring 'broader' measures. This could be done, for example, by: (i) supplementing the IMD income score with other deprivation measures that might pick up some broader, relevant causations; and / or (ii) using the overall IMD score, but where estimated values are used for Wales.

3.3.8 Average wholesale bill size

AVERAGE WHOLESALE BILL SIZE IS OUTSIDE OF A RETAILER'S EFFICIENT MANAGEMENT CONTROL. SEPARATELY, HOWEVER, OFWAT MAY WANT TO REFLECT A WHOLESALE EFFICIENCY CHALLENGE (AND WHOLESALE INFLATION) IN RETAIL COST ALLOWANCES PROJECTED OVER PR19.

Another driver of the bad debt element of retail costs is likely to be average wholesale bill size. This is because:

- for the same probability of customers going into arrears or default, the higher the average bill, the greater the £s loss impact on companies; and
- there may be some (secondary) relationship between the likelihood of a customer entering arrears or default and the average size of bill.

In addition, *from the perspective of a retailer*, the average size of wholesale bill is outside of efficient management control – because:

- if considered from the perspective of retailers being 'independent', clearly there is nothing they can directly do to influence the size of bill (i.e. wholesale charges are a pure input cost); and
- even to the extent that one considers retailers as part of integrated end-to-end water companies, the average size of wholesale bill cannot be influenced by them *for the duration of a price control.*

There is an important separate, but related, issue following from the above, regarding how Ofwat sets projected retail cost allowances over PR19. That is to say, clearly (integrated) companies can make efficiency savings and, relatedly, Ofwat will set an efficiency challenge to companies on the wholesale side. Similarly, the wholesale control will also reflect the impact of general inflation, given that Ofwat will index the

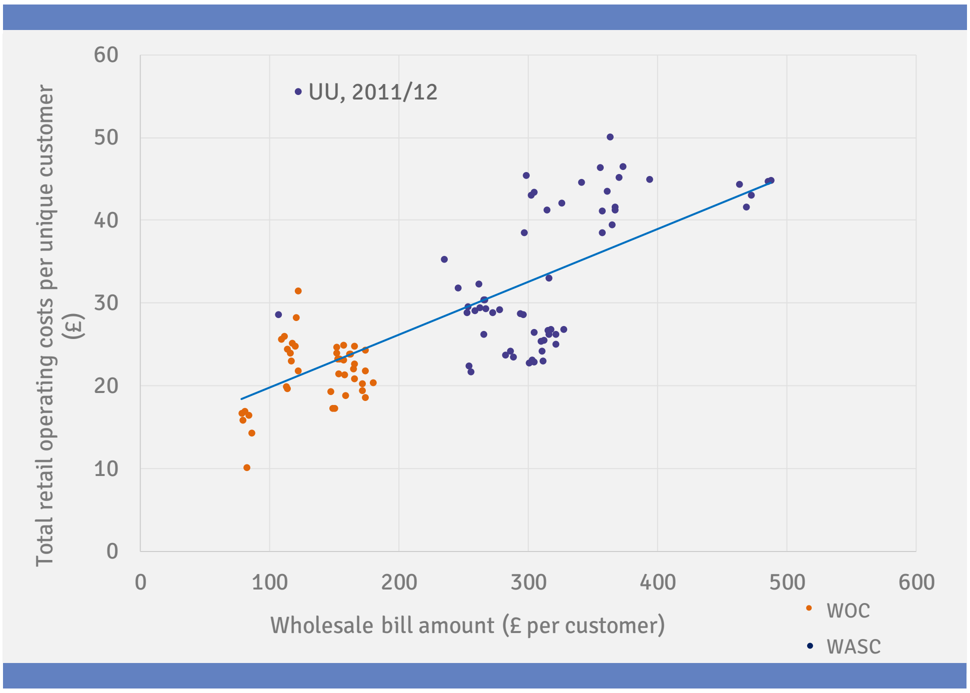
RCV by CPI(H). In that context, it may be appropriate for Ofwat to factor in a projected profile of wholesale costs when setting retail cost allowances on a forward-looking basis. This should not, however, be confused with the issue of retail cost controllability for the purpose of retail efficiency benchmarking.

Of relevance to the above, the 'in principle' connection between average (wholesale) bill size and retail costs was accepted at PR14. For example, in reviewing the econometric models used to support South West Water's bad debt related special factor cost claim, PWC (reviewing on behalf of Ofwat) stated: *"Overall… we considered that they provided evidence of the link between deprivation, **average bills** [emphasis added] and doubtful debt charges."*[15]

The following figure shows a scatterplot of average wholesale bill size (in £s) against total retail operating costs per unique customer. Our measure of wholesale bill size has been calculated by multiplying the total retail turnover with the percentage share of the wholesale bill size.

As expected, we find a strong positive correlation between unit costs and the size of wholesale bill.

Figure 21: Scatterplot of **total retail operating costs per unique customer** against **average wholesale bill size**



Source: Economic Insight

The following table provides a summary of the related correlation coefficients.

[15] *'PwC review of South West Water's doubtful debt cost models.' PWC letter to Ofwat (April 2014).*

Table 15: Correlation between costs and wholesale bill size

Measures		Wholesale bill (£ per customer)
Total operating cost	Total cost	0.436
Total operating cost	Unit cost	0.707
Bad debt	Total cost	0.476
Bad debt	Unit cost	0.797
Non-bad debt related retail costs	Total cost	0.372
Non-bad debt related retail costs	Unit cost	0.299

Source: Economic Insight

3.3.8.1 Key implications

- It is uncontentious that average wholesale bill size will significantly drive the bad debt element of household retail costs. This is supported by our descriptive statistics analysis, where we find strong, positive correlations, both in relation to total and unit measure of retail cost.
- We therefore find there is a strong case for including a measure of wholesale bill size within our generalised modelling framework (both for the total industry model – and in relation to the bad debt specific model).

3.3.9 Population transience

THERE ARE A NUMBER OF ‘IN PRINCIPLE’ WAYS IN WHICH POPULATION TRANSIENCE COULD IMPACT RETAIL COSTS.

Consistent with our approach of evaluating household retail cost drivers from first principles, we considered that ‘population transience’ might be a relevant factor. Here, by ‘transience’ we mean the propensity of a population to move into, or out of, a company supply area (movements within supply areas may also be relevant). We considered that this might impact retail related costs in the following ways:

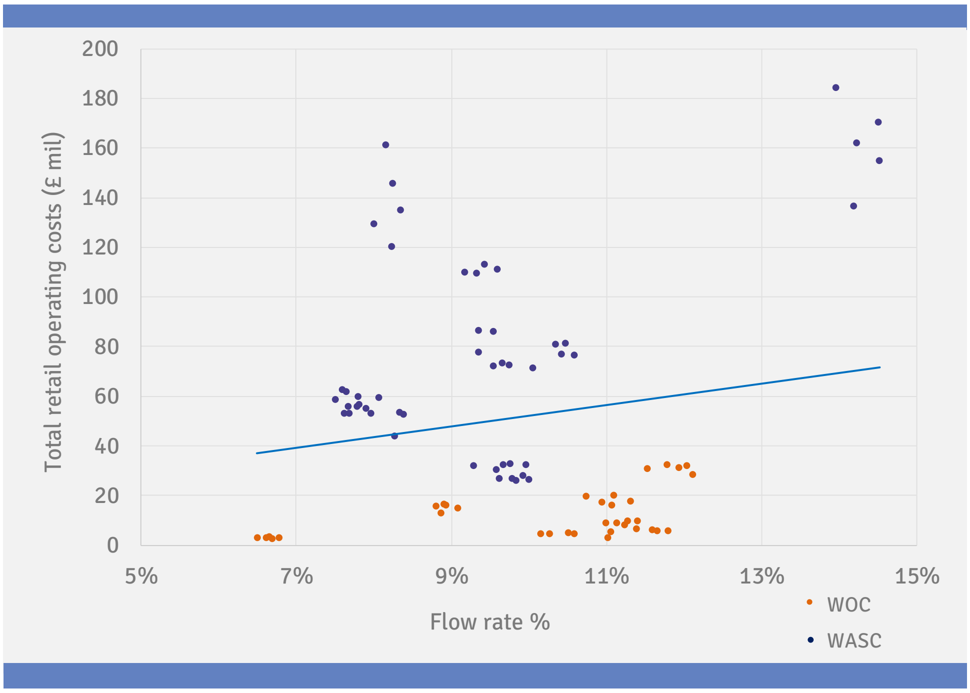
- Firstly, it could be related to the likelihood of customers falling into arrears or default, thereby increasing bad debt costs.
- Secondly, and relatedly, it might make debt management and recovery more difficult and expensive for companies (due to the need to ‘chase’ customers as they change address).
- Thirdly, and entirely separately from bad debt related costs, transience might impact account management costs. This is because as customers move into, or out of, a region, so companies will need to ‘set up’, ‘close’ or ‘transfer’ accounts for those customers.
- Fourthly, and finally, transience could impact meter reading costs. This could be the case if, for example, when customers move into, or out of, a region, they manually provide meter readings to companies, which allow companies to defer undertaking their own physical meter reading.

As population transience is demonstrably outside of efficient management control, it seemed appropriate for us to undertake an analysis of this. Our main measure of population transience is based on population inflow and outflow data, as published by the ONS. This data tracks movements between UK local authorities.

For summary purposes, the following two figures show:

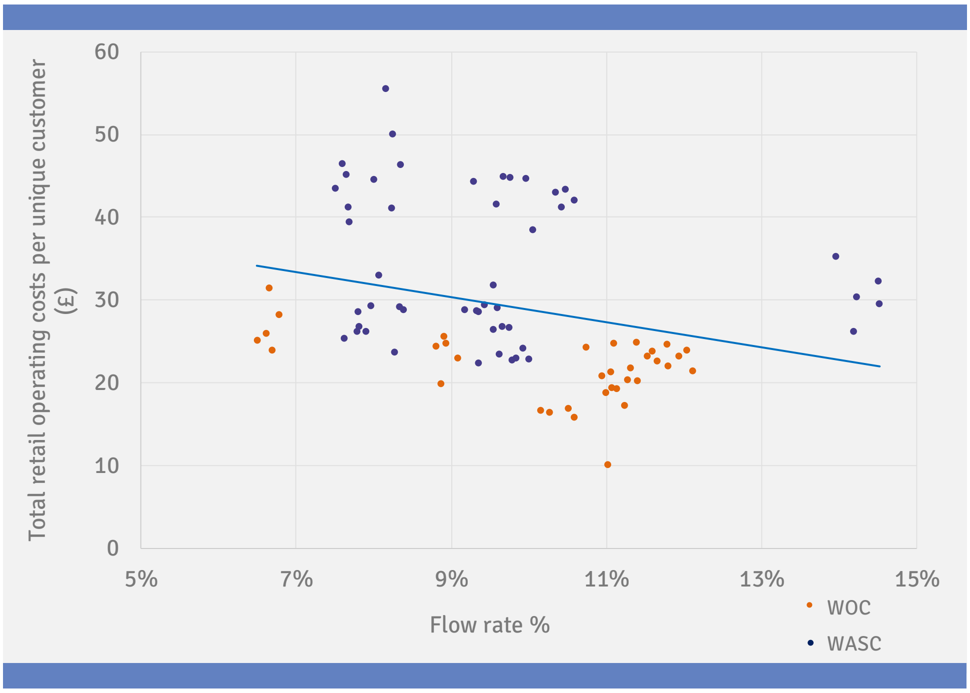
- A scatterplot of total retail operating costs against total population flow rate (i.e. the sum of absolute population flow divided by the population, where population is defined as the number of people that are residents in a region).
- A scatterplot of retail operating costs *per unique customer* against total population flow rate.

Figure 22: Scatterplot of **total retail operating costs** against **population flow rate (%)**



Source: Economic Insight

Figure 23: Scatterplot of **total retail operating costs per unique customer** against **population flow rate (%)**



Source: Economic Insight

Interestingly, while the first figure shows a *positive* relationship between the flow rate and absolute retail costs, the second suggests a *negative* relationship between flow rate and unit retail costs. The following tables summarise the associated correlation coefficients, for population flows expressed as numbers and as rates.

Table 16: Correlation between **absolute** transience flow measures and costs

Measures		Inflow (000)	Outflow (000)	Total (000)
Retail operating costs	Total cost	0.881	0.869	0.875
Retail operating costs	Unit cost	0.223	0.207	0.215
Bad debt related retail costs	Total cost	0.818	0.809	0.814
Bad debt related retail costs	Unit cost	0.293	0.275	0.284
Non-bad debt related retail costs	Total cost	0.891	0.876	0.884
Non-bad debt related retail costs	Unit cost	0.022	0.015	0.018

Source: Economic Insight

Table 17: Correlation between transience **rate** measures and costs

Measures		Inflow (%)	Outflow (%)	Total (%)
Retail operating costs	Total cost	0.097	0.233	0.169
Retail operating costs	Unit cost	**-0.301**	**-0.279**	**-0.294**
Bad debt related retail costs	Total cost	0.094	0.219	0.161
Bad debt related retail costs	Unit cost	**-0.222**	**-0.222**	**-0.225**
Non-bad debt related retail costs	Total cost	0.094	0.233	0.168
Non-bad debt related retail costs	Unit cost	**-0.326**	**-0.272**	**-0.302**

Source: Economic Insight

3.3.9.1 Key implications

We generally find population transience to be positively associated with total costs, but negatively associated with unit costs. In relation to the latter, this seems contrary to our hypothesis that increased transience might increase customers' propensity to fall into arrears or default, thereby increasing (for example) bad debt costs on a 'unit' basis. However, we should highlight the fact that this analysis excludes other key drivers of bad debt costs (socioeconomic performance and bill size) and so one cannot reach a strong conclusion one way or the other based on this alone. Indeed, as our econometric modelling subsequently establishes, once other factors are controlled for, we do find transience to be a statistically valid driver of debt-related costs.

The positive association between flow rate (%) and total (absolute) cost measures is consistent with increased population flow propensity driving up other, say account management, related costs (note, because we are measuring a flow rate, this should avoid us conflating transience with scale). Given the above findings, therefore, we consider it reasonable to include a measure of transience (specifically the flow rate) within our generalised modelling framework.

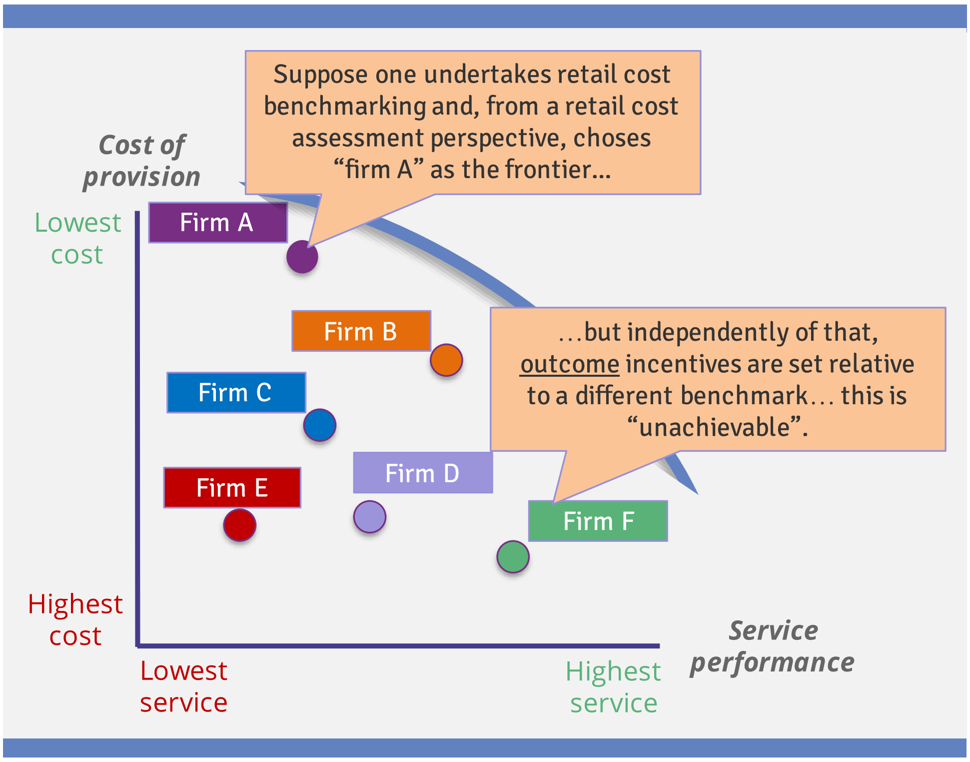
3.3.10 Retail service quality performance

AS A MATTER OF PRINCIPLE, COST ASSESSMENT SHOULD NOT BE UNDERTAKEN INDEPENDENTLY OF A CONSIDERATION OF SERVICE QUALITY.

In principle, in any industry one would expect service quality to be related to costs to a degree. Put simply, it costs more to provide a higher quality of service than a lower level of service.

Related to the above, there is a strong 'in principle' reason to consider service quality performance within any approach to cost assessment. That is, in economics, the 'frontier' should be both technically and allocatively efficient. Consequently, if one could perfectly identify the frontier efficient firm in the context of cost assessment, its quality performance should also be the respective benchmark for service. In simple terms, this means there are inherent conceptual difficulties with undertaking cost assessment and benchmarking *independently of an assessment of service quality performance* – as illustrated in the following figure.

Figure 24: Illustration of the 'cost efficiency' and 'service quality' interaction



Source: Economic Insight

Importantly, this point has been recognised by various regulators and competition authorities. For example:

- The econometric models Ofcom relies upon in benchmarking Royal Mail Group's cost efficiency explicitly include quality of service variables.[16]

- In its approach to efficiency benchmarking in relation to electricity distribution, Ofgem also includes quality measures. For example, in their related report to Ofgem, Frontier Economics state: *"it is necessary to take account of quality of supply in our totex benchmarking to ensure that the model is not biased, and the estimates of DNOs' efficiency reflect the quality of supply delivered."*[17]

- In its redetermination of Bristol Water's PR14 price control settlement, the Competition and Markets Authority raised concerns regarding the independent benchmarking of costs and outcomes.[18]

In relation to household retail, however, there is a practical problem. Namely, that there is not any particularly 'good' measure of retail service performance in our view. Specifically, the main available measures are:

- **The SIM.** This has been the historical measure, and incentive mechanism for, service quality in the sector. In relation to retail cost assessment, perhaps the main issue with SIM is that is covers both 'wholesale' and 'retail' issues (and is

[16] *For example, see Deloitte report for Ofcom 'Econometric benchmarking in the UK postal sector.' (May 2016).*
[17] *For example, see: 'Total cost benchmarking at RIIO-ED1 – Phase 2 report – Volume 1.' Frontier Economics on behalf of Ofgem (2013).*
[18] *'Bristol Water plc A reference under section 12(3)(a) of the Water Industry Act 1991 Report.' The CMA (2015). See page 283.*

arguably more wholesale driven). Further, even the more retail specific elements of it may be influenced by customers' perception of performance at the wholesale level. Ofwat intends to replace SIM at PR19 with a new incentive mechanism: WaterworkCX.

- **CCWater customer research measures.** CCWater publishes a range of measures regarding customers' perception of service in the water sector. Measures include: 'customer satisfaction'; 'value for money'; and 'Net Promoter Score.' The main challenge with these measures in the context of retail cost assessment are that they reflect customer perception of service in a broad sense, and so are hard to relate to tangible or specific service areas or activities. As a result, they may be driven primarily by service factors *unconnected with retail* (i.e. customers' perceptions may be more driven by wholesale performance).

'Previous analysis of retail costs has also found no intuitive or robust relationship with service quality.'

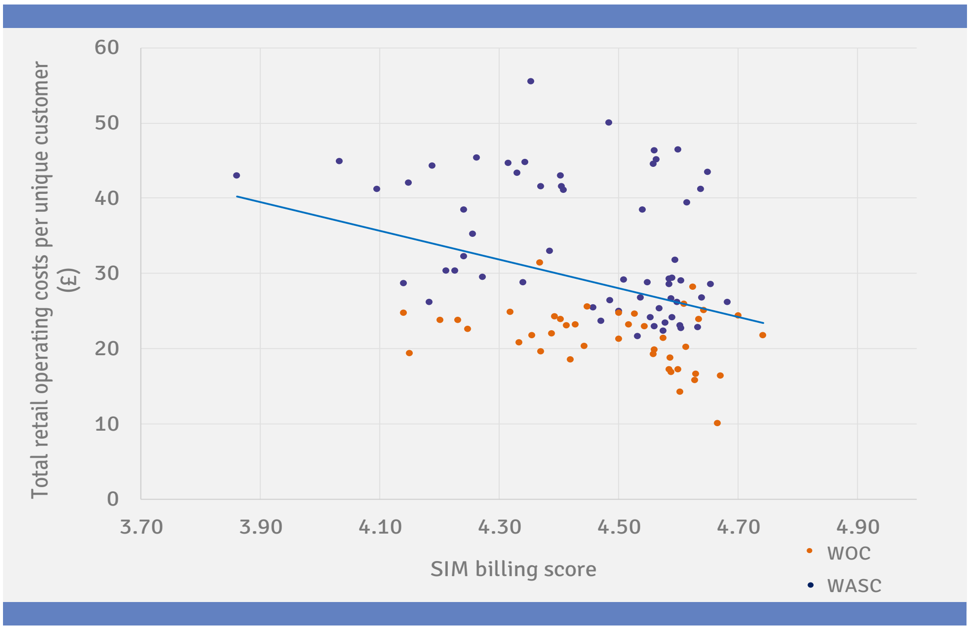
Consequently, in relation to the specific measures available to test, there is no expectation that these will necessarily reveal the logical interaction between retail service performance and cost. Indeed, we note that previous analysis of retail costs has also found no intuitive or robust relationship with service quality. For example, in our input price pressure work for Yorkshire Water at PR14, where we examined a range of retail econometric models, we also explored the use of SIM. Here, we found: *"With regard to the SIM score, we were unable to find any model specification including this variable was robust."*[19]

Notwithstanding the above, we have examined the relationship between the available measures and various cost metrics. Of the available measures, we think that SIM billing score is perhaps most relevant – and so, for summary purposes, the following scatterplots show:

- Total retail operating costs per unique customer against SIM billing score.
- Customer service related retail costs per unique customer against SIM billing score.

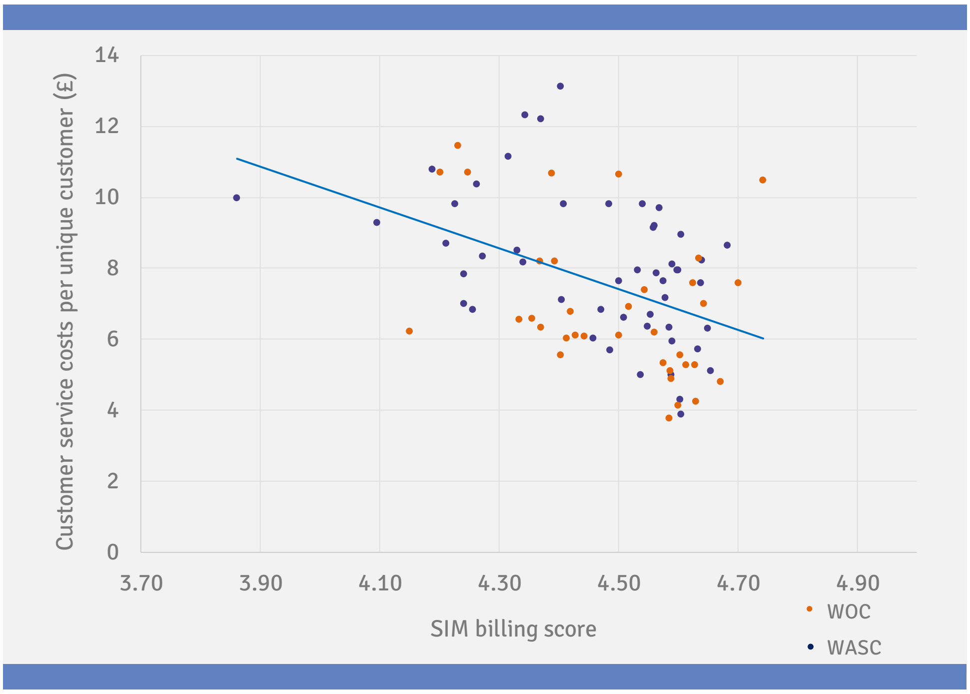
[19] *'Retail HH Input Price Pressure and benchmarking: a report for Yorkshire Water.' Economic Insight (2014).*

Figure 25: Scatterplot of **total retail operating costs per unique customer** against **SIM billing score**



Source: Economic Insight

Figure 26: Scatterplot of **customer service related retail costs per unique customer** against **SIM billing score**



Source: Economic Insight

Both of the above show a negative association between SIM and retail costs. The table overleaf sets out correlation coefficients relating to various combinations of cost and service quality measure.

Table 18: Correlation between retail service quality measures and retail operating costs

Measures		Overall SIM	SIM billing score	SIM billing above upper quartile	SIM billing above median	Customer satisfaction	Net promoter score
Total operating cost	Total cost	-0.275	-0.285	-0.262	-0.208	-0.349	-0.125
Total operating cost	Unit cost	-0.298	-0.348	-0.234	-0.314	-0.326	-0.122
Bad debt	Total cost	-0.251	-0.264	-0.249	-0.203	-0.308	-0.065
Bad debt	Unit cost	-0.174	-0.234	-0.183	-0.194	-0.195	-0.024
Non-bad debt related retail costs	Total cost	-0.282	-0.290	-0.260	-0.200	-0.368	-0.173
Non-bad debt related retail costs	Unit cost	-0.405	-0.420	-0.237	-0.406	-0.453	-0.266
Customer service cost	Total cost	-0.277	-0.279	-0.296	-0.154	-0.385	-0.179
Customer service cost	Unit cost	-0.291	-0.443	-0.357	-0.461	-0.528	-0.291

Source: Economic Insight

WE ALSO TESTED FOR 'NON-LINEARITY' IN THE COST / SIM RELATIONSHIP, BUT STILL FOUND A (COUNTER-INTUITIVE) NEGATIVE ASSOCIATION.

The above correlations are consistent with the selected scatterplots shown previously – i.e. they are generally counterintuitive, suggesting no 'logical' cost / service quality relationship.

We further considered that it might be possible that the cost / quality relationship was 'non-linear' or discrete, rather than being continuous. For example, it might be possible that any cost/quality relationship is only observable in relation to quality performance *above a certain threshold.* We therefore examined a number of ways of testing for this, including testing for 'upper quartile' and 'median' cut-offs in SIM performance. However, we again found a negative relationship.

3.3.10.1 Key implications

The 'in principle' need to consider service performance when undertaking cost assessment is well-established and is consistent with economic theory. However, specifically in relation to the available measures of service performance in the water sector, the expected relationships are not observed at a retail level. There are multiple potential explanations as to why the observed relationships are often 'negative' here. These include:

- The measures of 'quality' (i.e. SIM) are not sufficiently robust.
- Because retail is a sufficiently homogenous activity from the perspective of customers that **very large differences in 'quality' are needed to shift the dial in costs** – and so nothing is apparent in the data.
- Retail is sufficiently less efficient than wholesale, such that <u>some</u> (less efficient) companies can simultaneously improve cost efficiency and retail service performance. Consequently, cost / quality trade-offs are not apparent in the data. This might arise because the industry focus on retail performance is still relatively recent, following the introduction of separate controls for retail at PR14.
- **There may be legacy effects,** such that companies with a low SIM score today are incurring more operating costs in order to improve service, but there is a "lag" to see the service gain.

In relation to the above, we asked Drs Anthony and Karli Glass to consider the observed negative relationship. Their views were as follows:

"We are not surprised that modelling the quality-cost nexus is proving problematic. This is because accounting for service quality appropriately is a recurring issue in the academic efficiency and productivity literature. The problem relates to the adequacy of the data on service quality." They further commented: *"Interestingly, three companies in the data set with lower than average SIM scores over the 2011-2015 period (Southern Water, South West Water and Thames Water) are amongst the companies with reasonably high total operating costs, where Thames has the highest total operating cost in the sample. The negative coefficient on SIM in a cost model may therefore be due to the scale of the operation of the companies."*

In general, we cannot determine with certainty as to 'which' explanations primarily explain the observed negative relationship. Indeed, they are not mutually exclusive. Regardless, in practical terms this would seem to suggest that: (i) including such measures in an econometric modelling framework is unlikely to be successful; but similarly, (ii) there is no clear, credible basis on which one could make off-model adjustments, given that these would need to start from the same measures of retail performance.

In the long-term, the aspiration should be to better measure the retail elements of industry service performance, such that more robust analysis can be undertaken. Absent that, for PR19, perhaps the most appropriate approach is for retail service performance to be taken under consideration when selecting the 'frontier' for cost assessment. This is to avoid extremes, such that the 'challenge' for companies is either 'unduly hard' or 'unduly easy'.

Given the above, while we test service performance within our generalised framework because of its theoretical correctness, we consider it unlikely that this will prove suitable in any finalised models used to set cost benchmarks.

4. Econometric Modelling and Results

This chapter sets out our econometric cost benchmarking analysis for household retail. We first provide an overview of our approach, before summarising the key results of our modelling. We then set out diagnostic tests. In total, we generate a suite of 16 econometric models that can be used to generate efficiency gap estimates for the PR19 household retail control.

4.1 Our approach

We have developed our approach to address the key issues, as set out in the preceding chapter. The key features of our approach are as follows.

- **A combination of top-down total retail operating cost models, and separate bottom-up models** for bad debt and non-bad debt related retail operating costs**.** This is consistent with Ofwat's indicated methodology for PR19.
- We use a **general to specific approach.** This begins with 'generalised' models including the full set of potential cost drivers, informed by our descriptive statistical analysis set out above. We then eliminate variables one-by-one, depending on their statistical significance, to arrive at a set of specific models.
- In addition to specific models selected according to statistical significance, we develop **alternative models.** Reflecting the key issue of how best to balance 'statistical significance' against 'engineering intuition', we developed *alternative* versions of our total retail operating cost models, including additional explanatory variables.
- We used two approaches to addressing the panel structure of the dataset: (i) we estimated **pooled models** using OLS; and (ii) we estimated **random effects models**, using GLS.

- We included **scale** (customer numbers) and **scope** (dual versus single service customers) within the models in two ways: In **specification A** we included separate variables for dual and single service customers. In **specification B** we included a separate single service customer variable, alongside a 'total customers' variable.

In total, this generated a suite of 16 econometric models, as summarised in the table below. The remainder of this section discusses these issues in more detail.

Table 19: Suite of econometric cost models

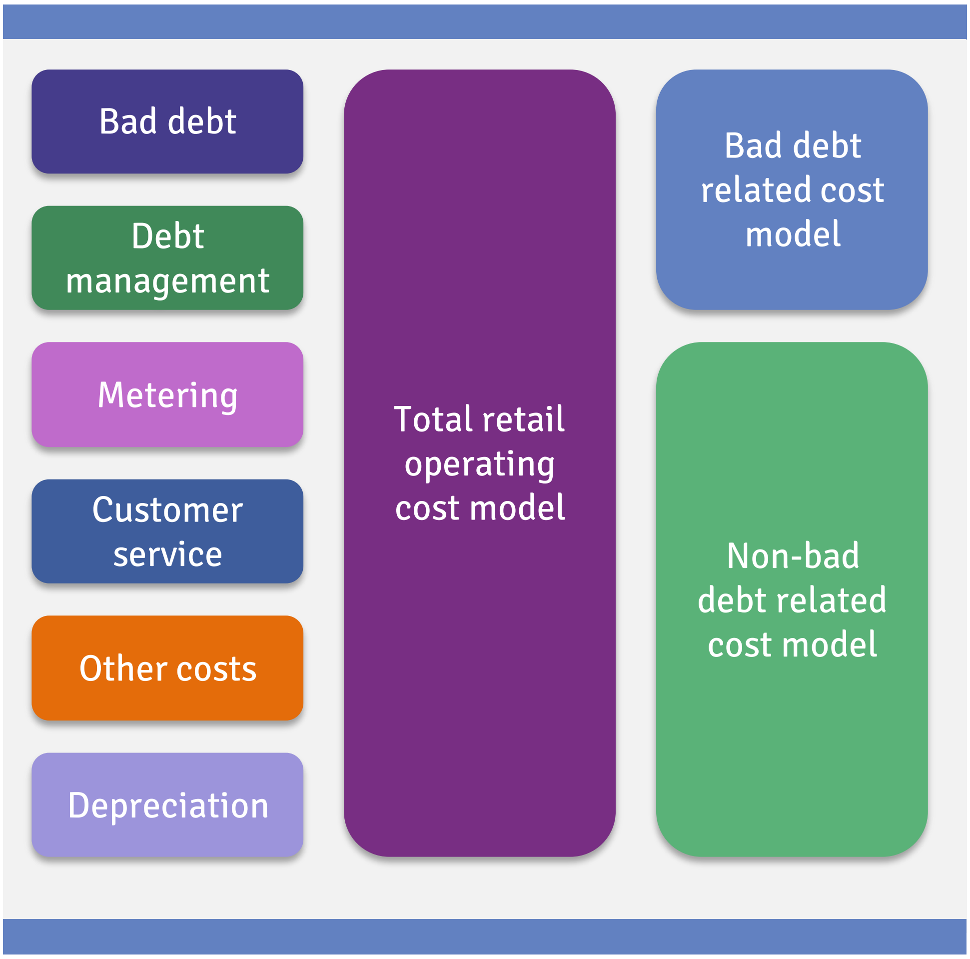
Model	Dependent variable	Panel structure	Estimation technique	General to specific approach	Approach to number of customers
A1	Total retail operating costs	Pooled	Ordinary Least Squares	Statistical significance	Separate dual and single service customer variables
A2	Bad debt related retail operating costs	Pooled	Ordinary Least Squares	Statistical significance	Separate dual and single service customer variables
A3	Non-bad debt related retail operating costs	Pooled	Ordinary Least Squares	Statistical significance	Separate dual and single service customer variables
A4	Total retail operating costs	Pooled	Ordinary Least Squares	Alternative approach	Separate dual and single service customer variables
A5	Total retail operating costs	Random effects	Generalised Least Squares	Statistical significance	Separate dual and single service customer variables
A6	Bad debt related retail operating costs	Random effects	Generalised Least Squares	Statistical significance	Separate dual and single service customer variables
A7	Non-bad debt related retail operating costs	Random effects	Generalised Least Squares	Statistical significance	Separate dual and single service customer variables
A8	Total retail operating costs	Random effects	Generalised Least Squares	Alternative approach	Separate dual and single service customer variables
B1	Total retail operating costs	Pooled	Ordinary Least Squares	Statistical significance	Total customers; single service customers
B2	Bad debt related retail operating costs	Pooled	Ordinary Least Squares	Statistical significance	Total customers; single service customers
B3	Non-bad debt related retail operating costs	Pooled	Ordinary Least Squares	Statistical significance	Total customers; single service customers
B4	Total retail operating costs	Pooled	Ordinary Least Squares	Alternative approach	Total customers; single service customers
B5	Total retail operating costs	Random effects	Generalised Least Squares	Statistical significance	Total customers; single service customers
B6	Bad debt related retail operating costs	Random effects	Generalised Least Squares	Statistical significance	Total customers; single service customers
B7	Non-bad debt related retail operating costs	Random effects	Generalised Least Squares	Statistical significance	Total customers; single service customers
B8	Total retail operating costs	Random effects	Generalised Least Squares	Alternative approach	Total customers; single service customers

Source: Economic Insight

4.1.1 Top-down and bottom-up models

In line with Ofwat's indicated methodology, we include both top-down models including the whole of retail operating costs (opex and capital costs) as the dependent variable – alongside bottom-up models that separately model: bad debt related costs (namely bad debt and debt management); and all other retail operating costs. This is summarised in the figure below.

Figure 27: Top-down and bottom-up cost models

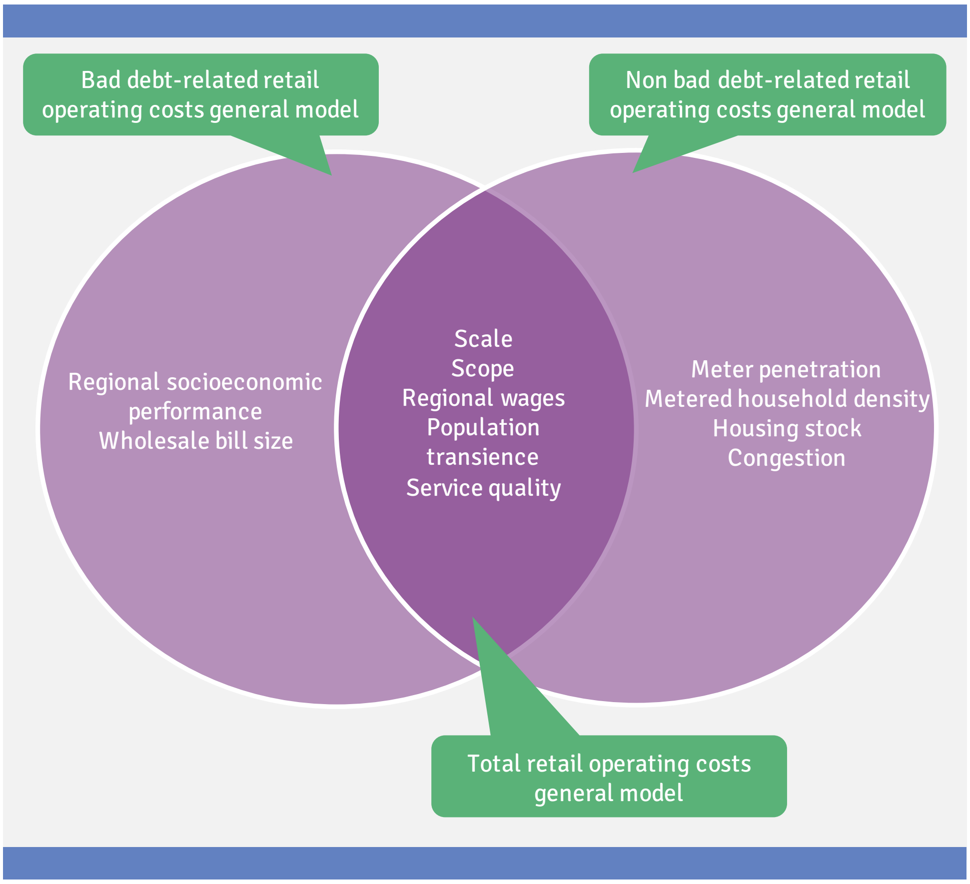


Source: Economic Insight

4.1.2 General to specific approach

In light of the evidence set out in the preceding chapter, for each of the dependent variables set out above, we developed a set of generalised econometric models that included the full range of potentially relevant cost drivers. The variables included in these models are summarised in the figure below.

Figure 28: Variables included in generalised econometric models



Source: Economic Insight

Starting from these generalised models, we next applied a general to specific approach. This began with a regression of the general, unrestricted model. We then eliminated the variables with the lowest statistical significance (i.e. highest p-value) on-by-one, continuing until the remaining variables had significance levels that approached 10%. In keeping with a 'liberal' approach to statistical significance, we retained variables with significance levels that approached 10% and retained the correct sign. We removed variables that were statistically significant, but which were incorrectly signed.

4.1.3 Alternative models

In addition to the liberal approach to statistical significance described above, we included alternative versions of the total retail operating cost models. This is in the spirit of achieving a 'balance' between statistical significance and engineering intuition. In practice, general to specific modelling for total retail costs results in explanatory variables that broadly correspond to scale (and scope) alongside drivers of bad debt related costs. Very few drivers that explain non-bad debt related costs are included in these models, reflecting bad debt's dominance within the overall retail cost stack. As such, having estimated general to specific total retail operating cost models, we then re-estimated the models including metering variables and retained any that were appropriately signed, even if they were not statistically significant.

4.1.4 Panel structure

There are a range of options for addressing the panel structure of the data. In our modelling we focused on pooled models, estimated using OLS and random effects models estimated using GLS. This is consistent with the approach to wholesale econometrics at PR14. Pooled and random effects models are also complementary, so there are advantages in including both approaches. Pooled OLS models allow the calculation of year-by-year efficiency scores, which random effects models do not. On the other hand, random effects models exploit the panel structure of the data, and distinguish between time-varying errors (assumed to include inefficiency) and statistical noise.

4.1.5 Treatment of scale and scope

WE USED TWO ALTERNATIVE APPROACHES TO THE TREATMENT OF 'SCALE' AND 'SCOPE' WITHIN THE MODELS.

As we set out in the preceding chapter, scale – the number of customers that companies serve – is the most important driver of household retail costs. It is closely related to scope – the extent to which companies serve dual (water and wastewater) versus single service (water-only or wastewater-only) customers. The importance of scope was recognised at PR14 in the setting of different cost allowances for dual and single service customers (alongside for metered versus unmetered customers).

In view of the importance of these two issues, and the extent of their interrelation, we used two distinct specifications to incorporate them within the econometric models.

- **Specification A** includes separate variables for the numbers of single and dual service customers.
- **Specification B** includes a variable for the total number of customers, alongside a separate variable for the number of single service customers.

We considered potential alternative approaches, such as including a company-level dual service dummy variable. Overall, we considered that our proposed specifications had most theoretical merit, while others had material downsides. For instance, a dual service dummy would be likely to capture overall WaSC versus WoC relative (in)efficiency, rather than the direct impact of scope.

'[W]e do not think that there are strong reasons to consider either approach more credible than the other.'

Overall, both specifications A and B have advantages and disadvantages and we do not think that there are strong reasons to consider either approach more credible than the other. As we show subsequently, they do result in materially different efficiency score estimates for *some* companies. As such, we consider implied efficiency gap estimates separately for the two 'model sets' based on specifications A and B.

We summarise the advantages and disadvantages of the two approaches in the table overleaf. With respect to model set A, we note the following.

- Model set A provides a flexible specification, and includes a larger number of relevant cost drivers, thereby reducing the risk that the models suffer from omitted variable bias.
- On the other hand, some companies have no dual service customers. As a result, the implied coefficient on dual service customers will represent an average impact on cost across all companies, including those that have no such customers.

- The marginal effects of the number of dual and single service customers are difficult to interpret.
- In practice, the efficiency gap estimates from model set A are very sensitive to the choice of benchmark.

Turning to model set B:

- Model set B has the advantage of a more parsimonious regression specification.
- Coefficients in model set B are easier to interpret, though the specification with respect to scope economies is less flexible.
- In practice, model set B includes fewer relevant cost drivers than model set A.
- Model set B appears to fit the data better. This is not necessarily an advantage, as one cannot distinguish, a priori, between inefficiency and issues related to model specification.

Table 20: Advantages and disadvantages of approaches to scope economies

Approach	**Model set A** Separate dual & single service customer variables	**Model set B** Separate total & single service customer variables
Advantages	More flexible specification Larger number of relevant cost drivers Reduced risk of omitted variable bias	More parsimonious regression specifications Estimated coefficients easier to interpret Appears to fit data better
Disadvantages	Many companies have no dual customers Marginal effects are difficult to interpret Gaps sensitive to benchmark choice	Less flexible specification Fewer relevant cost drivers included (potential omitted variable bias)

Source: Economic Insight

4.2 Dataset

4.2.1 Dataset summary

To undertake our modelling, we created a panel dataset, covering all of the companies in England and Wales and the time period 2011/12 to 2016/17. The following table provides a complete summary of our finalised dataset.

Table 21: Overview of our dataset

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
Total number of 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
Metered households 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
Sales-related occupations pay	Regional labour costs	ONS	2011/12-2016/17
IMD income score	Socioeconomic performance	ONS/Welsh Government	2010, 2015
Property repossessions	Socioeconomic performance	ONS	2011/12-2016/17
House price to income ratio	Socioeconomic performance	ONS	2011/12-2016/17
Wholesale bill size	Wholesale bill	Company data share	2011/12-2016/17
Population flow	Population transience	ONS	2011/12-2015/16
SIM billing score	Retail service quality	Ofwat	2011/12-2016/17

Source: Economic Insight

4.2.2 Data cleaning

To ensure the robustness of our dataset, our key steps were as follows.

- Percentage of flats – There was only one year's data for the percentage of flats in an area. As this is unlikely to change significantly over time, we assumed this value for the following years.

- IMD data – IMD values are only available for the years 2010 and 2015. We assumed 2015 values for the preceding years (note, alternative measures of socioeconomic performance, which are time variant, were also tested).

- Datashare - If data for a company was missing for any given year, then we replaced it by an average of the adjacent years. For example, in the datashare, no values were reported for Dee Valley and Sutton and East Surrey for the financial year 2013/14, so these have been replaced by an average of the corresponding values of 2012/13 and 2014/15.

- Cambridge Water – Cambridge Water became part of South Staffordshire Water in 2013; and so after this point their accounts are reported jointly. To keep this group consistent across the years, we have merged the values corresponding to Cambridge Water and South Staffordshire Water for the preceding years.

- Dual service customers – The log of zero values is a missing value in STATA. The variable number of dual service customers is zero for WoCs and creates many missing values when transformed into the log form. To circumvent this issue, we added one to this variable, giving us a zero value for the WoCs in the log form.

- Levels and percentages – Many of the variables sourced from the ONS are reported in absolute levels, rather than in percentage terms. To ensure that the regression captures more than the scale factor associated with these variables, we transformed them into rates by dividing each value with either the number of people living in that region, or by the number of households in that region, as appropriate.

- Congestion peak hours – Data for Wales were only available on an average basis, rather than a peak hour basis. Data for Wales were estimated by applying the average 'peaking factor' for England (i.e. the proportion by which peak hours traffic speeds exceed average speed) to average speed for Wales.

- We dropped values from 2011/12. Cost values for a number of companies appeared materially different to other years, which generated concern given the potential that this reflected accounting separation issues.

4.3 General to specific modelling

This section summarises the general to specific modelling, detailing the order in which variables were eliminated, alongside the reasons for which they were eliminated – that is to say, whether this was due to statistical significance, or the coefficient having the wrong sign. As models A4, A8, B4 and B8 used the alternative approach described above, they are omitted from this section.

4.3.1 Model set A

The table below summarises general to specific modelling for models A1 to A3.

Table 22: General to specific modelling for models A1 to A3

Variable	Model A1 Total costs (ln)			Model A2 Bad debt costs (ln)			A3 Non- bad debt costs (ln)		
	Included	Order eliminated	Reason	Included	Order eliminated	Reason	Included	Order eliminated	Reason
Ln(single service customers)	✓	-	-	✓	-	-	✓	-	-
Ln(dual service customers)	✓	-	-	✓	-	-	✓	-	-
Metered households (%)	✗	5	Significance	-	-	-	✓	-	-
Metered households to mains length	✗	6	Significance	-	-	-	✓	-	-
Flats (%)	✓	-	-	-	-	-	✗	2	Significance
Ln(traffic speed)	✗	1	Significance	-	-	-	✓		-
Ln(sales-related pay)	✗	2	Significance	✗	1	Significance	✗	1	Significance
IMD income (%)	✓	-	-	✓	-	-	-	-	-
House price to income (ratio)	✗	8	Wrong sign	✗	4	Significance	-	-	-
Property repossessions (%)	✗	3	Significance	✗	3	Significance	-	-	-
Ln(wholesale bill)	✓	-	-	✓	-	-	-	-	-
Population flow (%)	✗	9	Significance	✓	-	-	✗	4	Wrong sign
SIM billing score (%)	✗	7	Wrong sign	✗	5	Wrong sign	✗	5	Wrong sign
Time trend	✗	4	Significance	✗	2	Significance	✗	3	Significance

Source: Economic Insight

The table below summarises general to specific modelling for models A5 to A7.

Table 23: General to specific modelling for models A5 to A7

Variable	Model A5 Total costs (ln)			Model A6 Bad debt costs (ln)			A7 Non- bad debt costs (ln)		
	Included	Order eliminated	Reason	Included	Order eliminated	Reason	Included	Order eliminated	Reason
Ln(single service customers)	✓	-	-	✓	-	-	✓	-	-
Ln(dual service customers)	✓	-	-	✓	-	-	✓	-	-
Metered households (%)	✗	1	Significance	-	-	-	✓	-	-
Metered households to mains length	✗	3	Significance	-	-	-	✗	4	Significance
Flats (%)	✗	7	Significance	-	-	-	✗	1	Significance
Ln(traffic speed)	✗	4	Significance	-	-	-	✓	-	-
Ln(sales-related pay)	✗	8	Significance	✗	1	Significance	✗	2	Significance
IMD income (%)	✓	-	-	✓	-	-	-	-	-
House price to income (ratio)	✗	6	Significance	✗	3	Significance	-	-	-
Property repossessions (%)	✓	-	-	✗	5	Significance	-	-	-
Ln(wholesale bill)	✓	-	-	✓	-	-	-	-	-
Population flow (%)	✗	2	Significance	✗	2	Significance	✗	5	Significance
SIM billing score (%)	✗	9	Wrong sign	✗	6	Wrong sign	✗	3	Significance
Time trend	✗	5	Significance	✗	4	Significance	✓	-	-

Source: Economic Insight

4.3.2 Model set B

The table below summarises general to specific modelling for models B1 to B3.

Table 24: General to specific modelling for models B1 to B3

Variable	Model B1 Total costs (ln)			Model B2 Bad debt costs (ln)			B3 Non- bad debt costs (ln)		
	Included	Order eliminated	Reason	Included	Order eliminated	Reason	Included	Order eliminated	Reason
Ln(total customers)	✓	-	-	✓	-	-	✓	-	-
Ln(single service customers)	✘	10	Significance	✘	2	Significance	✓	-	-
Metered households (%)	✘	1	Significance	-	-	-	✓	-	-
Metered households to mains length	✘	9	Wrong sign	-	-	-	✘	1	Significance
Flats (%)	✘	8	Significance	-	-	-	✘	3	Significance
Ln(traffic speed)	✘	7	Wrong sign	-	-	-	✓	-	-
Ln(sales-related pay)	✘	3	Significance	✘	3	Significance	✘	2	Wrong sign
IMD income (%)	✓	-	-	✓	-	-	-	-	-
House price to income (ratio)	✘	2	Significance	✘	5	Significance	-	-	-
Property repossessions (%)	✓	-	-	✘	6	Significance	-	-	-
Ln(wholesale bill)	✓	-	-	✓	-	-	-	-	-
Population flow (%)	✘	4	Significance	✘	4	Significance	✘	5	Significance
SIM billing score (%)	✘	6	Wrong sign	✘	7	Wrong sign	✘	4	Wrong sign
Time trend	✘	5	Significance	✘	1	Significance	✘	6	Significance

Source: Economic Insight

The table below summarises general to specific modelling for models B5 to B7.

Table 25: General to specific modelling for models B5 to B7

Variable	Model B5 Total costs (ln)			Model B6 Bad debt costs (ln)			B7 Non- bad debt costs (ln)		
	Included	Order eliminated	Reason	Included	Order eliminated	Reason		Order eliminated	Reason
Ln(total customers)	✓	-	-	✓	-	-	✓	-	-
Ln(single service customers)	✓	-	-	✗	3	Significance	✓	-	-
Metered households (%)	✗	2	Significance	-	-	-	✓	-	-
Metered households to mains length	✗	10	Wrong sign	-	-	-	✗	5	Wrong sign
Flats (%)	✗	8	Significance	-	-	-	✗	4	Significance
Ln(traffic speed)	✗	7	Wrong sign	-	-	-	✓	-	-
Ln(sales-related pay)	✗	3	Significance	✗	4	Significance	✗	1	Significance
IMD income (%)	✗	9	Significance	✓	-	-	-	-	-
House price to income (ratio)	✗	4	Significance	✗	6	Significance	-	-	-
Property repossessions (%)	✓	-	-	✓	-	-	-	-	-
Ln(wholesale bill)	✓	-	-	✓	-	-	-	-	-
Population flow (%)	✗	1	Significance	✗	2	Significance	✗	3	Significance
SIM billing score (%)	✗	6	Wrong sign	✗	5	Wrong sign	✗	2	Significance
Time trend	✗	5	Significance	✗	1	Significance	✓	-	-

Source: Economic Insight

4.4 Results

4.4.1 Model set A

The table below presents the pooled OLS models from model set A (models A1 to A4), which include separate dual and single service customer variables. These models have the following functional forms.

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

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

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

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

Table 26: Model set A – pooled OLS models

Variables	Model A1 Total costs (ln)	Model A2 Bad debt costs (ln)	Model A3 Non-bad debt costs (ln)	Model A4 Total costs (ln)
Single service customers (ln)	0.536*** (0.000)	0.535*** (0.000)	0.498*** (0.000)	0.563*** (0.0000)
Dual service customers (ln)	0.122*** (0.000)	0.121*** (0.000)	0.263*** (0.000)	0.159*** (0.0000)
Metered households (%)			0.0143*** (0.0002)	0.00723* (0.062)
Metered to mains length (%)			-0.0155*** (0.001)	-0.00662** (0.041)
Proportion flats (%)	0.0571*** (0.000)			0.0604*** (0.001)
Peak traffic speed (ln)			-1.830*** (0.0000)	-0.364 (0.290)
IMD income (%)	0.164*** (0.000)	0.189*** (0.000)		0.155*** (0.000)
Wholesale bill	1.206*** (0.000)	1.744*** (0.000)		0.999*** (0.000)
Total internal migration (%)		0.0909*** (0.001)		
Constant	-10.02*** (0.000)	-14.37*** (0.000)	4.539*** (0.000)	-8.063*** (0.000)
Observations	89	89	89	89
R-squared (adjusted)	0.9284	0.9333	0.8743	0.9283

*Source: Economic Insight, ***p<0.01, ** p<0.05, * p<0.1*

The table below presents the random effects models from model set A (models A5 to A8). These models have the following functional forms.

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

Table 27: Model set A – random effects models

Variables	Model A5 Total costs (ln)	Model A6 Bad debt costs (ln)	Model A7 Non-bad debt costs (ln)	Model A8 Total costs (ln)
Single service customers (ln)	0.349*** (0.001)	0.532*** (0.000)	0.268** (0.025)	0.318*** (0.003)
Dual service customers (ln)	0.226*** (0.000)	0.184*** (0.003)	0.250*** (0.000)	0.246*** (0.000)
Metered households (%)			0.00214 (0.610)	0.00198 (0.500)
Proportion flats (%)				0.0526 (0.144)
Peak traffic speed (%)			-1.217** (0.047)	
IMD income (%)	0.0657 (0.167)	0.136*** (0.008)		0.105* (0.056)
Property repossessions (%)	0.107*** (0.000)			0.119*** (0.002)
Wholesale bill (ln)	0.341 (0.000)	1.235*** (0.002)		0.301 (0.213)
Trend			-0.0372** (0.014)	
Constant	-2.741 (0.103)	-10.25*** (0.000)	4.104* (0.067)	-3.836** (0.039)
Observations	89	89	89	89
R-squared (overall)	0.8957	0.9260	0.8539	0.9060

*Source: Economic Insight, ***p<0.01, ** p<0.05, * p<0.1*

4.4.2 Model set B

The table below presents the pooled OLS models from model set B (models B1 to B4), which include separate total and single service customer variables. These models have the following functional forms.

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

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

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

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

Table 28: Model set B – pooled OLS models

Variables	Model B1 Total costs (ln)	Model B2 Bad debt costs (ln)	Model B3 Non-bad debt costs (ln)	Model B4 Total costs (ln)
Total customers (ln)	0.877*** (0.000)	0.979*** (0.000)	1.061*** (0.000)	0.966*** (0.000)
Single service customers (ln)			-0.120*** (0.000)	-0.0690* (0.087)
Metered households (%)			0.00452*** (0.004)	0.00473*** (0.005)
Peak speed (ln)			-0.257* (0.062)	
IMD income (%)	0.0273*** (0.001)	0.0668*** (0.000)		0.0274*** (0.003)
Property repossessions (%)	0.121*** (0.000)			0.147*** (0.000)
Wholesale bill (ln)	0.659*** (0.000)	1.091*** (0.000)		0.480*** (0.000)
Constant	-6.974*** (0.000)	-11.31*** (0.000)	-3.200*** (0.000)	-6.502*** (0.0000)
Observations	89	89	89	89
R-squared (adjusted)	0.9821	0.9616	0.9676	0.9835

*Source: Economic Insight, ***p<0.01, ** p<0.05, * p<0.1*

The table below presents the random effects models from model set B (models B5 to B8). These models have the following functional forms.

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

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

Table 29: Model set B – random effects models

Variables	Model B5 Total costs (ln)	Model B6 Bad debt costs (ln)	Model B7 Non-bad debt costs (ln)	Model B8 Total costs (ln)
Total customers (ln)	1.043*** (0.000)	0.933*** (0.000)	1.069*** (0.000)	1.065*** (0.000)
Single service customers (ln)	-0.134** (0.041)		-0.138** (0.021)	-0.150** (0.030)
Metered households (%)			0.00461 (0.114)	0.00201 (0.400)
Peak speed (ln)			-0.327 (0.286)	
IMD income (%)		0.0553* (0.071)		
Property repossessions (%)	0.113*** (0.000)	0.147** (0.015)		0.130*** (0.000)
Wholesale bill (ln)	0.400*** (0.004)	1.165*** (0.000)		0.351** (0.019)
Time trend			-0.0349*** (0.002)	
Constant	-5.519*** (0.000)	-11.57*** (0.000)	-2.820** (0.011)	-5.446*** (0.000)
Observations	89	89	89	89
R-squared (overall)	0.9815	0.9639	0.9709	0.9824

*Source: Economic Insight, ***p<0.01, ** p<0.05, * p<0.1*

4.5 Summary diagnostic testing

Having developed a set of 16 econometric models, we subjected them to a range of diagnostic tests. Additional details of these are set out in the appendix. The table overleaf provides a summary of key results. We note the following.

- Variance inflation factors suggest that none of the models are subject to concerns over multicollinearity.
- In general, time dummies are not significant with the models (except for model A6).
- Breusch-Pagan tests consistently suggest that random effects models are preferred to pooled OLS models. In practice, we expect this result is driven by (relatively time invariant) inefficiency being 'mistaken' for firm-level effects. As such, we do not think that this finding, in itself, is a strong reason to 'prefer' random effects models.
- Tests indicate the potential for heteroscedasticity in models A4 and B4. This is of limited concern, as we use robust standard errors.
- There is some evidence of serial correlation in the random effects total operating cost models. This appears to be driven by the final year of data (2016/17). Coefficient estimates remain unbiased, although their efficiency may be reduced.

Table 30: Summary of diagnostic testing results

Model	Variance inflation factors (multicollinearity)	Time dummy joint significance	Breusch-Pagan test (random effects)	Breusch -Pagan/ Cook-Weisberg (heteroscedasticity)	Serial correlation
A1	Low concern	Time dummies not significant	Random effects preferred	Cannot reject null of homoscedasticity	-
A2	Low concern	Time dummies not significant	Random effects preferred	Cannot reject null of homoscedasticity	-
A3	Low concern	Time dummies not significant	Random effects preferred	Cannot reject null of homoscedasticity	-
A4	Low concern	Time dummies not significant	Random effects preferred	Reject null of homoscedasticity	-
A5	Low concern	Time dummies not significant	Random effects preferred	-	Reject null of no serial correlation
A6	Low concern	Time dummies significant	Random effects preferred	-	Cannot reject null of no serial correlation
A7	Low concern	Time dummies not significant	Random effects preferred	-	Cannot reject null of no serial correlation
A8	Low concern	Time dummies not significant	Random effects preferred	-	Reject null of no serial correlation
B1	Low concern	Time dummies not significant	Random effects preferred	Cannot reject null of homoscedasticity	-
B2	Low concern	Time dummies not significant	Random effects preferred	Cannot reject null of homoscedasticity	-
B3	Low concern	Time dummies not significant	Random effects preferred	Cannot reject null of homoscedasticity	-
B4	Low concern	Time dummies not significant	Random effects preferred	Reject null of homoscedasticity	-
B5	Low concern	Time dummies not significant	Random effects preferred	-	Reject null of no serial correlation
B6	Low concern	Time dummies not significant	Random effects preferred	-	Cannot reject null of no serial correlation
B7	Low concern	Time dummies not significant	Random effects preferred	-	Cannot reject null of no serial correlation
B8	Low concern	Time dummies not significant	Random effects preferred	-	Reject null of no serial correlation

Source: Economic Insight

5. Efficiency Gap Estimates

This chapter sets out our estimates of firm-level efficiency gaps, based on the suite of econometric models described in the preceding sections. We first discuss the key issues in moving from econometric model results to efficiency gap estimates. Then, considering these issues 'in the round', we arrive at a set of underlying assumptions for the calculation of efficiency gaps – under three scenarios. Finally, using these scenarios, we derive the implied efficiency gaps.

5.1 Issues in the calculation of efficiency gaps

CALCULATING EFFICIENCY GAPS FROM ECONOMETRIC COST MODELS IS COMPLICATED BY UNCERTAINTY OVER IDENTIFYING THE EFFICIENCY FRONTIER AND OVER HOW MUCH OF ANY GAP CAN BE CLOSED, AND HOW QUICKLY.

Econometric cost models provide evidence of the relationship between costs and their drivers *at average levels of efficiency*. However, inefficiency is not itself a distinct component within the models; and instead is captured within model residuals. This leads to two difficulties in generating efficiency gap estimates from econometric models.

- The efficiency frontier cannot be perfectly identified, and residuals cannot be assumed to consist wholly of inefficiency.
- It is uncertain as to "how much" of any efficiency gap can be closed, and "how quickly" this can be done.

5.1.1 Uncertainty over the efficiency frontier

Although it is common practice to infer efficiency challenges from model residuals (particularly in the context of economic regulation), one cannot assume that the whole of any residual represents inefficiency. Residuals comprise a combination of inefficiency, random noise, and errors in the regression specification. In fact, the academic literature suggests that residuals may largely capture issues that are *unrelated* to inefficiency. This limitation needs to be accounted for when calculating efficiency gaps. Similarly, one therefore needs to be careful that any benchmark chosen is not in some way 'atypical', even after such adjustments.

These points are widely acknowledged by both academics and regulators. Relevant regulatory references to this issue include the following.

- *"In keeping with the view that there is some noise in the data that the COLS approach fails to account for, we have decided to apply what we would view to be a conservative 25% noise adjustment – our assumption is that, on average, 25% of an IM"s deviation from the frontier can be attributed to noise. In doing this we recognise that as the split between inefficiency and noise is unobserved, any adjustment is necessarily to some extent a matter of judgement."* – ORR (2013).[20]
- Ofwat applied residual adjustments at PR04 and PR09 – in its PR09 method, Ofwat stated: *"[we will] continue to expect each company to catch up 60% of the difference from the benchmark company over five years; [and we will] make an adjustment as we did at PR04 to residuals to take some account of the potential for errors*." – Ofwat (2008).[21] Note, at PR14, the decision to adopt an 'upper quartile' benchmark can be thought of as analogous to setting a 'frontier' and then reducing residuals (i.e. they are alternative ways of addressing the same issue – as explicitly acknowledged by Ofgem, below).
- "*We have benchmarked the efficient level of totex for each DNO using the upper quartile … rather than the frontier to allow for other factors that may influence the DNOs' costs".*[22] Ofgem (2013).

Relevant references from academic literature include the following.

- "*The disturbance arises for several reasons, primarily because we cannot hope to capture every influence on an economic variable in a model, no matter how elaborate… there are many other contributors to the disturbance in an empirical model. Probably the most significant is errors of measurement".*[23] – Greene (2003).
- *"COLS residuals comprises both inefficiency and noise"* – Street (2003).[24]
- *"The appropriate approach to the assessment estimate of efficiency is to scale down the residuals in the existing models".*[25] – Cubbin (2004).
- *"OLS and COLS have a major weakness because residuals in the estimation reflect a combination of relative efficiency, measurement error in the dependent variable, and statistical noise, rather than inefficiency only. As a result, the final point-estimates of 'efficiency' should be discounted before using them in formulating price policies for regulatory purposes".*[26] – Chung (2011).

5.1.2 Uncertainty over how much of the gap can be closed, and how quickly

Regulators need to decide **how much** of any gap it is feasible for an inefficient company to close; and **how quickly** an inefficient company can do so. Relatedly, regulators have financeability duties that require them to ensure that (efficient)

20 PR13 Efficiency Benchmarking of Network Rail using LICB. ORR (2013).
21 Setting price limits for 2010-15: Framework and approach. Ofwat (2008).
22 RIIO-ED1 business plan expenditure assessment – methodology and results. Ofgem (2013).
23 Econometric Analysis. 7th ed. William .H. Greene (2003).
24 How much confidence should we place in efficiency estimates? Andrew Street. (2003).
25 Assessing Ofwat's Efficiency Econometrics: A report for Water UK. John Cubbin (2004).
26 Review of building energy-use performance benchmarking methodologies. William Chung (2011).

companies are able to finance the proper running of their operations, which they need to take account of before making efficiency targets excessively demanding. Further, as regulators must consider customer welfare in both the near and long-term, it is important to consider trade-offs. Setting an efficiency challenge that requires a gap to be closed 'immediately' might encourage short-term 'cost cutting' at the expense of long-run welfare. This is pertinent to retail, where the only realistic way to achieve very large cost savings in the short term would be to reduce headcount.

5.1.3 Policy tools

POLICY TOOLS ARE AVAILABLE TO ADDRESS THESE ISSUES, BUT NEED TO BE CONSIDERED 'IN THE ROUND' AND IN THE CONTEXT OF THE ACTUAL MODEL RESULTS.

To address these issues, regulators have a number of 'policy tools' available to them.

- To address the problem that only a proportion of any residual represents inefficiency, regulators can adjust residuals downward. This can use fixed percentage adjustments, or more advanced statistical approaches, such as stochastic frontier analysis. Regulators may also reflect this uncertainty in their choice of benchmark. For example, options include: (i) applying adjustments such as outlier removal and turnover rules, to exclude atypical observations; and (ii) selecting 'less demanding' benchmars than the absolute frontier (i.e. minimum residuals), such as upper quartile.

- To ensure efficiency targets are feasible, regulators can make percentage adjustments to estimated efficiency gaps. For instance, they could decide that, say, 60% of total inefficiency could feasibly be eliminated within the course of the price control. Relatedly, they could allow glide paths to efficient costs, dividing estimated efficiency gaps by the number of years in the price control.

These options are summarised in the table below.

Table 31: Key issues and relevant policy tools

Issue	Parameter	Policy tools
Frontier is not observable, so only a proportion of residuals represent inefficiency	Residuals adjustment	Percentage residual adjustments; statistical approaches (e.g. stochastic frontier).
	Frontier Selection	Upper quartile, upper quintile or average performance benchmark; pragmatic turnover rules; outlier treatment.
Uncertainty over how much and how quickly efficiency gap can be closed	Feasibility adjustment	Percentage adjustments to total efficiency gap.
	Glide path	Divide estimated gap by number of years in control.

Source: Economic Insight

Overall, the most important consideration in choosing between these options is that the eventual package of tools makes sense 'in the round'; and is considered within the context of actual model results.

5.2 Efficiency assumptions

For benchmarking PR19 household retail, we have developed three scenarios – low, central and high – with underlying assumptions, to generate efficiency gap estimates. The aim here is to set out transparently how different assumptions can affect the implied efficiency challenge, by generating a plausible range of options. This will allow stakeholders to consider for themselves which approach is most appropriate. The assumptions for each scenario are set out in the table below.

Table 32: Underlying assumptions for efficiency gap calculation

Parameter	Low case	Central case	High case
Model weights	Equal weights	Equal weights	Equal weights
Residual adjustment	None	None	None
Benchmark	Average	Upper quartile	Upper quintile
Glide path	Five-year	None	None

Source: Economic Insight

In triangulating across models, we need to make a practical decision as to how to deal with companies that perform better than the selected benchmark. Without making an adjustment, such firms would have *negative* efficiency gaps, so in practice one should perform an adjustment to ensure that overall efficiency gap estimates are zero at the lowest. One has a choice as to whether such an adjustment is made at the model level, or whether this is done *after* the weights are applied to the models. We have applied this adjustment after weights have been applied, as applying the adjustment at the model level can lead to some counterintuitive results, as firms receive a heavy 'penalty' for performing below the benchmark on a single model, even if they perform well ahead of the benchmark on all other models.

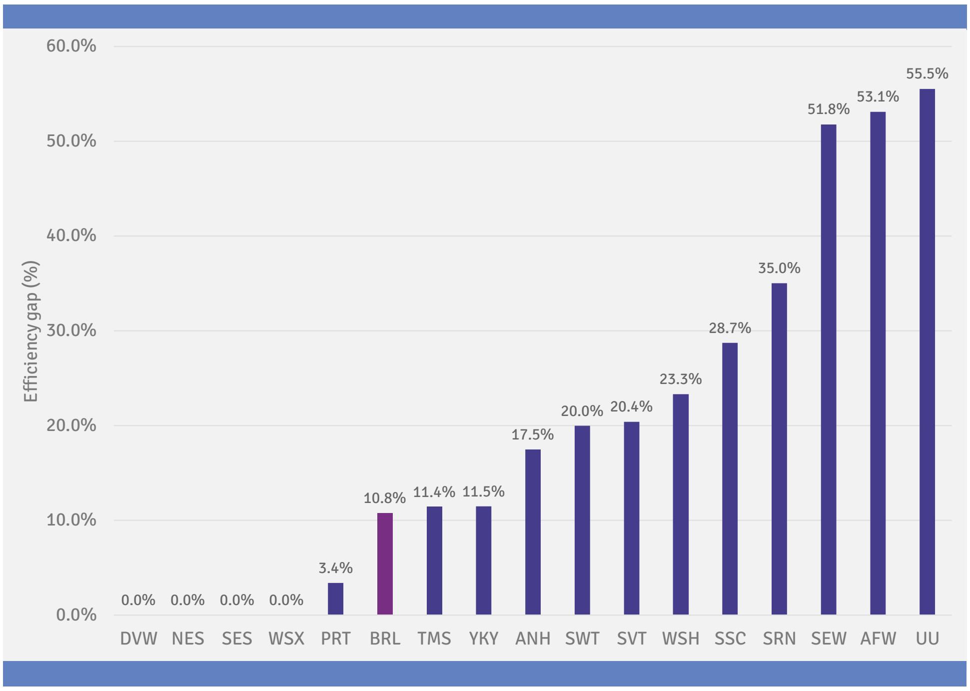
Following from the above, in the next section we set out the efficiency scores (i.e. % gaps to the benchmark) implied by our scenarios.

5.3 Efficiency scores

5.3.1 Central case

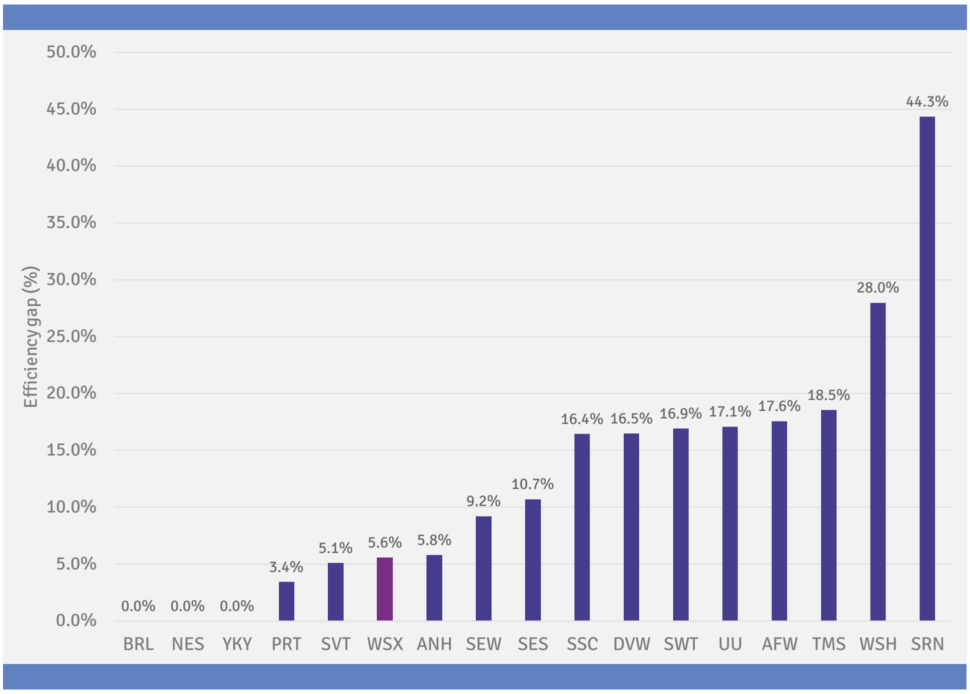
The figures below show estimates of the total efficiency gaps using the assumptions specified in the central case, for model sets A and B respectively.

Figure 29: Total eficiency gap estimates – model set A, central case



Source: Economic Insight

Figure 30: Total eficiency gap estimates – model set B, central case



Source: Economic Insight

The table below presents the total efficiency challenges for model sets A and B, alongside the average efficiency challenges across the two model sets. As our central case assumes no glide path, the assumption is that the entirety of these gaps would need to be closed in year 1 of PR19.

Table 33: Total efficiency challenges – central case

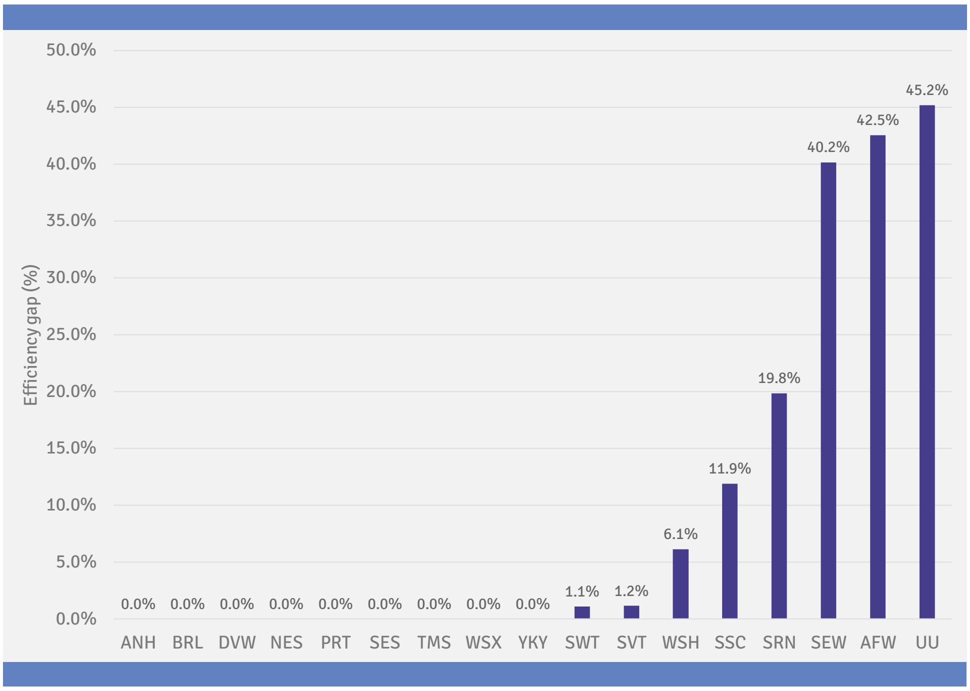
Company	**Total efficiency challenge**		
	Model set A	*Model set B*	*Average*
AFW	53.1%	17.6%	35.3%
ANH	17.5%	5.8%	11.6%
BRL	**10.8%**	**0.0%**	**5.4%**
DVW	0.0%	16.5%	8.2%
NES	0.0%	0.0%	0.0%
PRT	3.4%	3.4%	3.4%
SES	0.0%	10.7%	5.4%
SEW	51.8%	9.2%	30.5%
SRN	35.0%	44.3%	39.7%
SSC	28.7%	16.4%	22.6%
SVT	20.4%	5.1%	12.8%
SWT	20.0%	16.9%	18.4%
TMS	11.4%	18.5%	15.0%
UU	55.5%	17.1%	36.3%
WSH	23.3%	28.0%	25.6%
WSX	**0.0%**	**5.6%**	**2.8%**
YKY	11.5%	0.0%	5.7%

Source: Economic Insight

5.3.2 Low case

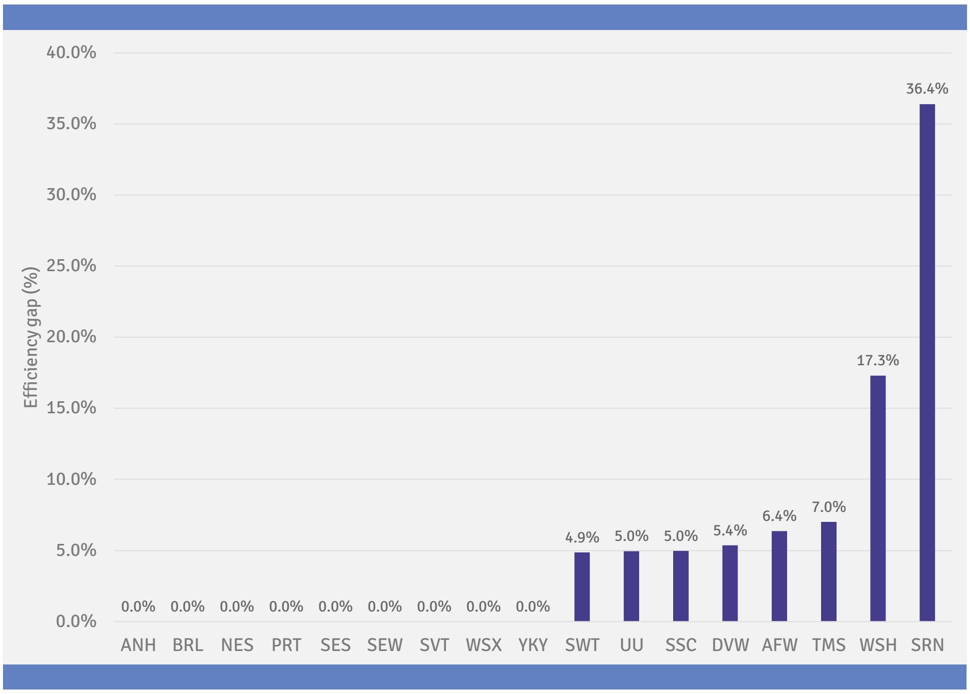
The figures below show estimates of the total efficiency gaps using the assumptions specified in the low (less challenging) case, for model sets A and B respectively.

Figure 31: Total eficiency gap estimates – model set A, low case (less challenging)



Source: Economic Insight

Figure 32: Total eficiency gap estimates – model set B, low case (less challenging)



Source: Economic Insight

The table below presents the total efficiency challenges for model sets A and B, alongside the average efficiency challenges across the two model sets. As our low case includes a glide path, under this approach the total efficiency gaps would be spread equally over the 5 years of PR19.

Table 34: Total efficiency challenges – low case

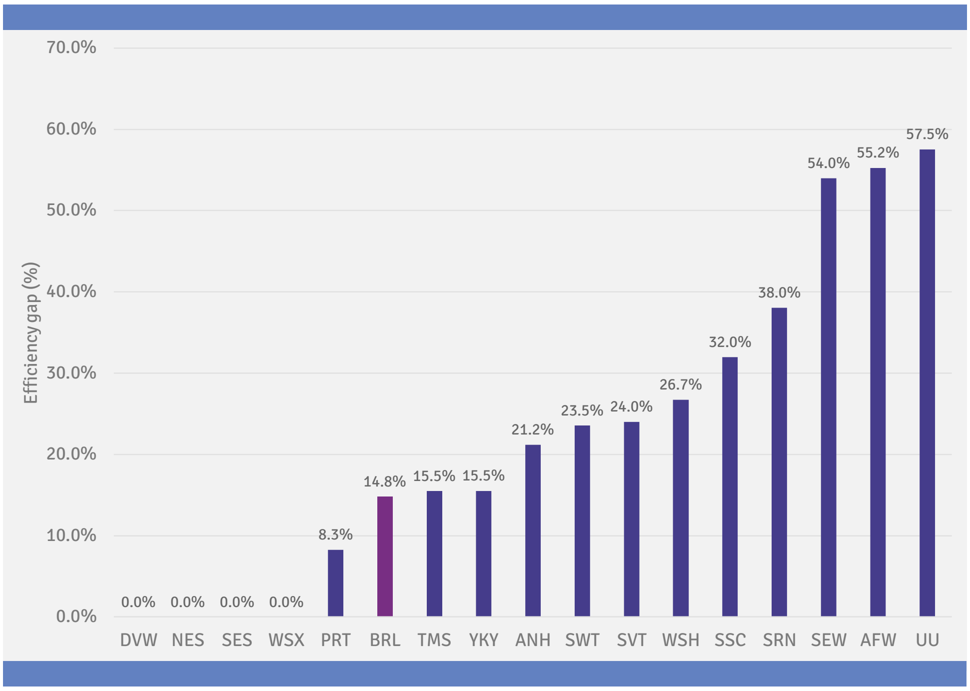
Company	Total efficiency challenge		
	Model set A	*Model set B*	*Average*
AFW	42.5%	6.4%	24.5%
ANH	0.0%	0.0%	0.0%
BRL	**0.0%**	**0.0%**	**0.0%**
DVW	0.0%	5.4%	2.7%
NES	0.0%	0.0%	0.0%
PRT	0.0%	0.0%	0.0%
SES	0.0%	0.0%	0.0%
SEW	40.2%	0.0%	20.1%
SRN	19.8%	36.4%	28.1%
SSC	11.9%	5.0%	8.4%
SVT	1.2%	0.0%	0.6%
SWT	1.1%	4.9%	3.0%
TMS	0.0%	7.0%	3.5%
UU	45.2%	5.0%	25.1%
WSH	6.1%	17.3%	11.7%
WSX	**0.0%**	**0.0%**	**0.0%**
YKY	0.0%	0.0%	0.0%

Source: Economic Insight

5.3.3 High case

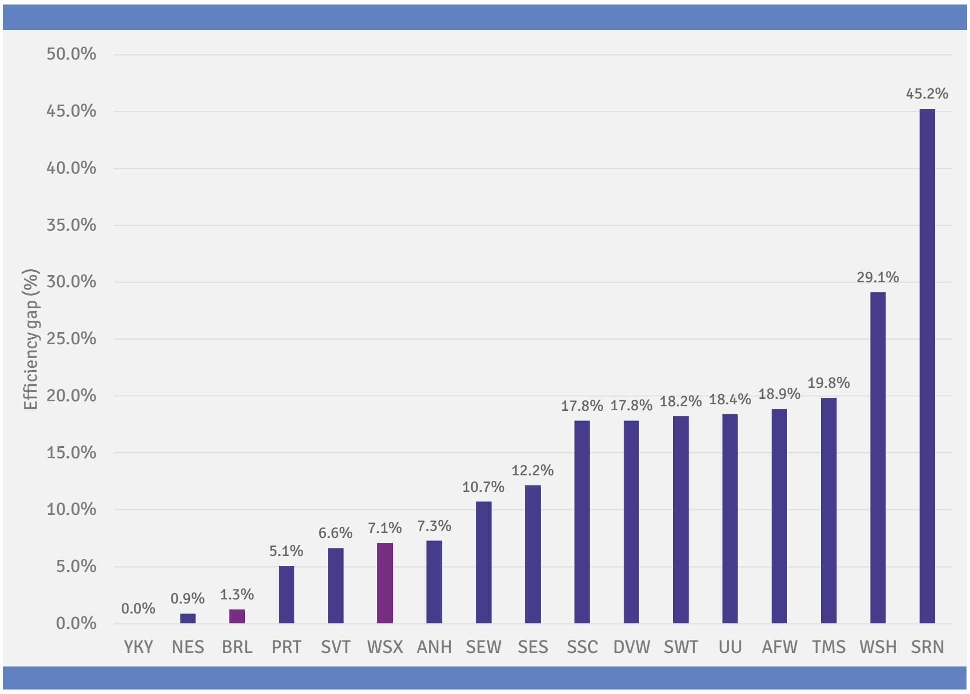
The figures below show estimates of the total efficiency gaps using the assumptions specified in the high (more challenging) case, for model sets A and B respectively.

Figure 33: Total eficiency gap estimates – model set A, high case (more challenging)



Source: Economic Insight

Figure 34: Total eficiency gap estimates – model set B, high case (more challenging)



Source: Economic Insight

The table below presents the total efficiency challenges for model sets A and B, alongside the average efficiency challenges across the two model sets. As the high case does not include a glide path, the presumption is that the entirety of these gaps would be closed in year 1 of PR19 under this scenario.

Table 35: Total efficiency challenges – high case

Company	Total efficiency challenge		
	Model set A	*Model set B*	*Average*
AFW	55.2%	18.9%	37.1%
ANH	21.2%	7.3%	14.2%
BRL	**14.8%**	**1.3%**	**8.0%**
DVW	0.0%	17.8%	8.9%
NES	0.0%	0.9%	0.4%
PRT	8.3%	5.1%	6.7%
SES	0.0%	12.2%	6.1%
SEW	54.0%	10.7%	32.3%
SRN	38.0%	45.2%	41.6%
SSC	32.0%	17.8%	24.9%
SVT	24.0%	6.6%	15.3%
SWT	23.5%	18.2%	20.9%
TMS	15.5%	19.8%	17.7%
UU	57.5%	18.4%	38.0%
WSH	26.7%	29.1%	27.9%
WSX	**0.0%**	**7.1%**	**3.5%**
YKY	15.5%	0.0%	7.8%

Source: Economic Insight

5.4 Wider sector comparators

As we explained in Chapter 2, in its methodology for PR19, Ofwat signalled that it will consider benchmarks outside the water sector when considering the appropriate efficiency challenge.

We have several concerns of the validity of such comparisons. Most obviously, there is no reason to suppose that the "efficient" level of cost, or quality of service, should be the same across different markets.

Nonetheless, for completeness, we have calculated an ACTS in relation to both:

- mobile virtual network operators (MVNOs); and
- energy retail (both in total, and the "controllable" element).

This reflects our separate analysis for Bristol and Wessex regarding the setting of retail margins, which found these industries to be 'most similar' to household retail in the water sector.

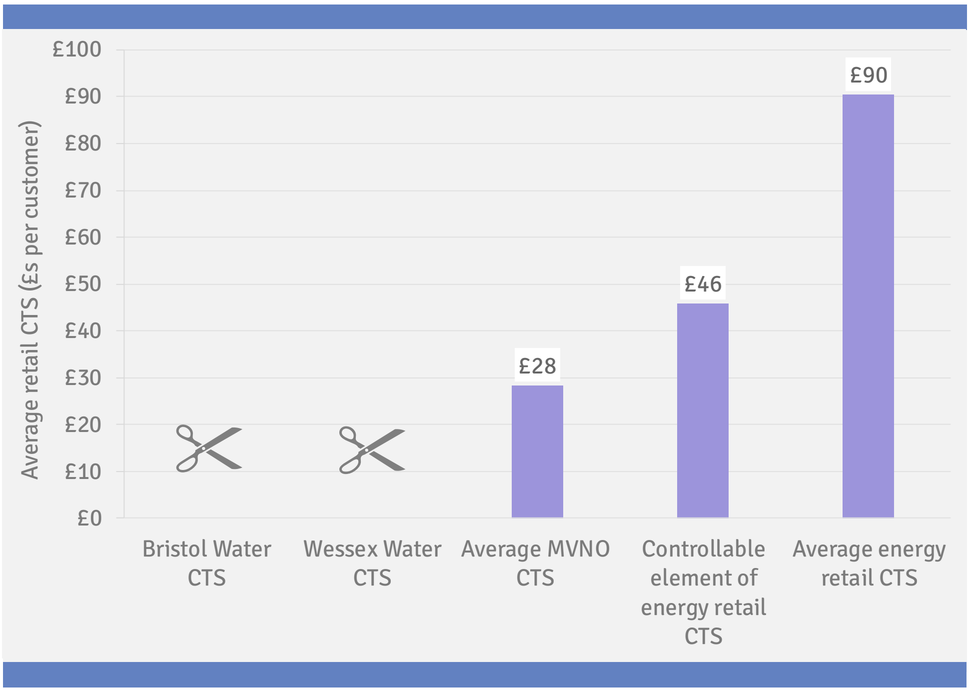
The key steps in calculating these cost benchmarks were as follows:

- **In relation to MVNOs,** total operating cost data (the numerator) was sourced for company statutory accounts, as per our margin analysis. Only 'indirect' costs (i.e. excluding costs of goods sold, which relate to the purchase of bandwidth wholesale) were included, to ensure that we identified only the *'retail'* element of costs. Customer numbers were calculated from: (i) the combined market share for MVNOs; and (ii) total mobile connection numbers, as published by Ofcom.[27] The operators included in this analysis are: Tesco Mobile, Virgin Mobile, Lycamobile, and Lebara.

- **In relation to energy**, our total retail cost data is based on the Consolidated Segmented Statements of the 'big six' energy retailers: (EDF, Centrica, Eon, RWE Npower, Scottish Power and SSE) using 2016 data. Again, to ensure that we only included retail (and household) related costs, the precise figures we have used relate to 'indirect costs' for domestic customers only. Our customer numbers are based on data published by Ofgem in its Retail Energy Markets report, which relates to 2016.[28]

We adjusted the accounting data for inflation, so that the implied average cost numbers are comparable to those shown for the water sector. Our results are summarised in the following figure.

Figure 35: Comparison of average retail cost to serve across sectors



Source: Economic Insight

Key points to note are:

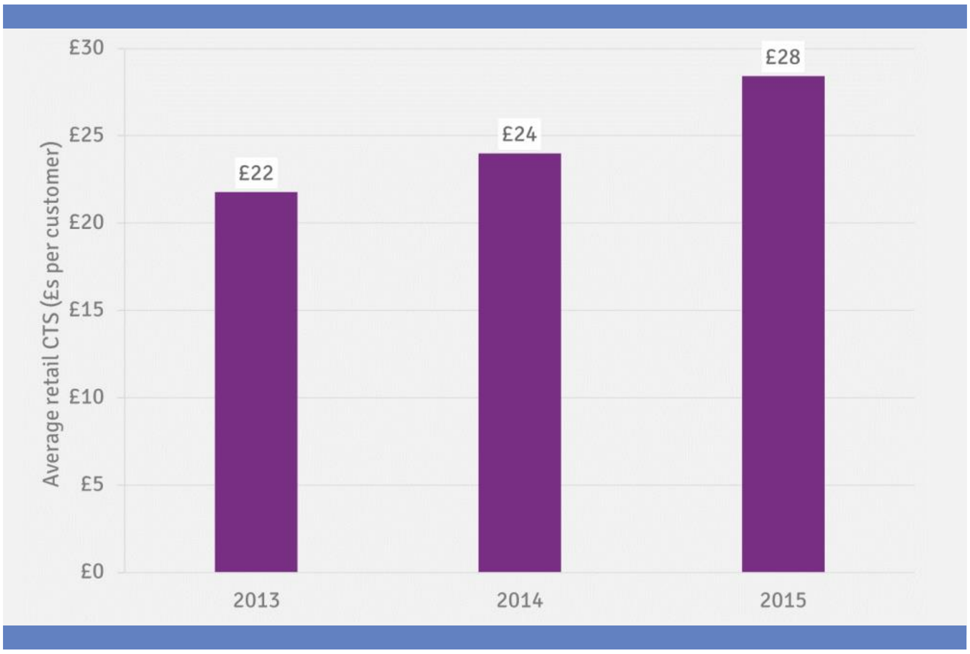
- The average cost to serve for energy retailers is higher than for both Bristol and Wessex, making energy retailers an unhelpful comparator. This is unsurprising given: (i) the average size of energy bills (£1,142)[29] is substantially higher than

[27] *'Communications Market Report 2016.'. Ofcom (2016).*
[28] *'Retail Energy Markets in 2016'. Ofgem (2016).*
[29] *https://www.ofgem.gov.uk/publications-and-updates/infographic-bills-prices-and-profits*

that for water at £398;[30] and (ii) retail costs are a higher proportion of the total value chain in energy than in water.

- The cost to serve of MVNOs overall is closer to that of Bristol and Wessex. Previous analysis suggested the average CTS for MVNOs tended to be slightly below that for water retail, and we additionally calculated MVNO average CTS over time – as shown in the following figure.

Figure 36. MVNOs retail CTS over time



Source: Economic Insight

The above figure shows that the average CTS has increased by 30% in the three years to 2015. Consequently, this explains why the average CTS for MVNOs is now *slightly above* that of water retail, relative to analysis undertaken for earlier years, indicating that MVNO CTS was *slightly below* that of water retail. Interestingly, of course, this also shows how even firms in highly competitive retail service markets can face significant inflationary cost pressures over time.

[30] http://www.water.org.uk/news-water-uk/latest-news/household-water-and-sewerage-bills-2016-17

6. Conclusions and Recommendations

This final chapter of our report sets out our conclusions and recommendations. Here key points are that – overall, our analysis shows that valid econometric models can be developed, and so is supportive of Ofwat's proposal to use statistical benchmarking to set retail costs at PR19. In addition, our work further highlights the need for careful consideration of the balance between statistical validity and engineering intuition in a retail setting.

6.1 Conclusions

Our main conclusions arising from the analysis set out in this report are as follows:

- **It is possible to identify econometric benchmarking models for household retail that perform well on measures of statistical robustness and are intuitively sound.** Given this, the analysis contained here, collectively, is supportive of the use of econometric modelling for the purpose of setting efficient household retail costs at PR19.

- Related to the above**, our modelling identifies a range of key cost drivers which are reasonably outside of efficient management control,** for inclusion in benchmarking. Key cost drivers include:

  - measures of single and dual serve customers;
  - meter penetration;
  - socioeconomic factors (e.g. IMD); and
  - average wholesale bills size.

- **Regional wages are not found to be either an intuitively sound, nor statistically valid, driver of household retail costs.** Consequently, we do not think there is a case for them being included within any econometric cost assessment model.

- **Because of the predominance of 'scale' and 'bad debt' related cost drivers – there is a risk of overlooking other, intuitively sensible, drivers of cost.** Indeed, our descriptive statistics and modelling analysis identified several factors, which hitherto have received relatively little attention in a retail cost assessment context. These include, for example:
  - congestion;
  - housing stock; and
  - population transience.
- **The way in which scale and scope are accounted for within the econometric models can have a significant impact on implied efficiency scores for some companies.** While some companies have similar efficiency scores under both of our approaches to measuring customer numbers, efficiency scores for certain companies vary more materially across the two methods. Given that both approaches are analytically sensible, this issue requires careful consideration.
- **It is important not to conflate benchmarking with setting a future profile of allowed costs over time –** where in the latter, one may very well wish to take a case of 'foreseeable' changes that ultimately will impact retail costs. For example, wholesale bill size is 'out of efficient retail management control' and therefore, should be controlled for when undertaking benchmarking analysis. However, when setting a forward view of allowed costs, it would seem to be legitimate for Ofwat to take into account known changes in wholesale bill size – reflecting, for example: (i) the wholesale efficiency challenge Ofwat sets (which, all else equal, will *reduce* required retail costs); and (ii) general inflation allowed for at the wholesale level (which, all else equal, will *increase* required retail costs).

6.2 Recommendations

Following from the above, our recommendations are as follows:

- Consistent with Ofwat's draft proposals, **an econometric approach should be used to set allowed costs for household retail at PR19.**
- **We recommend that Ofwat pays particular attention to the wider range of potentially valid explanatory variables** (outside of efficient management control) which might be 'crowded out' by the predominance of 'scale' and 'bad debt' related drivers. We further suggest that the precise way in which such factors might impact costs should be evaluated with care – as the historical focus on issues such as bad debt and deprivation has meant these issues have not been considered in detail to date.
- **When converting econometric results to efficiency challenges, the assumptions should be considered holistically** and the resultant efficiency gaps 'sense checked', to ensure they are plausible and defendable.
- **Particular care needs to be taken when defining the frontier**; and there are dangers in being overly reliant on the performance of any one company. This does not necessarily mean that, for example, upper quartile performance is the 'right' answer. However, care and consideration should be given to the sensitivity of results to different definitions of the frontier.

- **Service quality should be a consideration when setting the frontier.** Our analysis found significant difficulties in incorporating service quality within cost models. For example, raw data suggest a negative correlation, while none of the models we tested including quality variables produced signs that accorded with our priors. As a matter of principle, however, at the frontier there must be a cost-quality trade-off. Therefore, our view is that this issue is best addressed at PR19 by the regulator paying attention to the quality performance of any firms identified as being a candidate for the ‘frontier’, to avoid setting unduly challenging or unduly lenient targets.

7. Appendix: Diagnostic Tests

Table 36: Summary of diagnostic testing results

Model	Breusch-Pagan (random effects)	Breusch-Pagan / Cook-Weisberg (heteroscedasticity)	Multicollinearity (Variance inflation factor)	Time dummy joint significance	Serial correlation
A1	p-value: 0.000 *Random effects preferred*	p-value: 0.064 *Cannot reject null of homoscedasticity*	Mean: 3.51 Max: 6.98 *Low concern*	p-value: 0.895 *Time dummies not significant*	-
A2	p-value: 0.000 *Random effects preferred*	p-value: 0.852 *Cannot reject null of homoscedasticity*	Mean: 3.55 Max: 6.78 *Low concern*	p-value: 0.784 *Time dummies not significant*	-
A3	p-value: 0.000 *Random effects preferred*	p-value: 0.593 *Cannot reject null of homoscedasticity*	Mean: 1.94 Max: 2.83 *Low concern*	p-value: 0.645 *Time dummies not significant*	-
A4	p-value: 0.000 *Random effects preferred*	p-value: 0.038 *Reject null of homoscedasticity*	Mean: 5.96 Max: 13.49 *Low concern*	p-value: 0.732 *Time dummies not significant*	-
A5	p-value: 0.000 *Random effects preferred*	-	Mean: 3.49 Max: 6.78 *Low concern*	p-value: 0.189 *Time dummies not significant*	p-value: 0.005 *Reject the null of no serial correlation*
A6	p-value: 0.000 *Random effects preferred*	-	Mean: 3.80 Max: 6.78 *Low concern*	p-value: 0.01 *Time dummies significant*	p-value: 0.357 *Cannot reject the null of no serial correlation*
A7	p-value: 0.000 *Random effects preferred*	-	Mean: 1.20 Max: 1.40 *Low concern*	p-value: 0.108 *Time dummies not significant*	p-value: 0.407 *Cannot reject the null of no serial correlation*
A8	p-value: 0.000 *Random effects preferred*	-	Mean: 4.11 Max: 8.12 *Low concern*	p-value: 0.160 *Time dummies not significant*	p-value: 0.005 *Reject the null of no serial correlation*
B1	p-value: 0.000 *Random effects preferred*	p-value: 0.082 *Cannot reject null of homoscedasticity*	Mean: 1.96 Max: 2.62 *Low concern*	p-value: 0.873 *Time dummies not significant*	-
B2	p-value: 0.000 *Random effects preferred*	p-value: 0.736 *Cannot reject null of homoscedasticity*	Mean: 1.71 Max: 2.07 *Low concern*	p-value: 0.492 *Time dummies not significant*	-

Model	Breusch-Pagan (random effects)	Breusch-Pagan / Cook-Weisberg (heteroscedasticity)	Multicollinearity (Variance inflation factor)	Time dummy joint significance	Serial correlation
B3	p-value: 0.000 *Random effects preferred*	p-value: 0.550 *Cannot reject null of homoscedasticity*	Mean: 1.33 Max: 1.44 *Low concern*	p-value: 0.112 *Time dummies not significant*	-
B4	p-value: 0.000 *Random effects preferred*	p-value: 0.018 *Reject null of homoscedasticity*	Mean: 4.48 Max: 9.81 *Low concern*	p-value: 0.799 *Time dummies not significant*	-
B5	p-value: 0.000 *Random effects preferred*	-	Mean: 3.52 Max: 5.79 *Low concern*	p-value: 0.186 *Time dummies not significant*	p-value: 0.007 *Reject the null of no serial correlation*
B6	p-value: 0.000 *Random effects preferred*	-	Mean: 1.99 Max: 2.62 *Low concern*	p-value: 0.082 *Time dummies not significant*	p-value: 0.352 *Cannot reject the null of no serial correlation*
B7	p-value: 0.000 *Random effects preferred*	-	Mean: 1.34 Max: 1.44 *Low concern*	p-value: 0.138 *Time dummies not significant*	p-value: 0.356 *Cannot reject the null of no serial correlation*
B8	p-value: 0.000 *Random effects preferred*	-	Mean: 4.16 Max: 7.81 *Low concern*	p-value: 0.132 *Time dummies not significant*	p-value: 0.008 *Reject the null of no serial correlation*

Source: Economic Insight

WE MAKE ECONOMICS RELEVANT

Economic Insight Limited

125 Old Broad Street
London
EC2N 1AR
0207 100 3746
www.economic-insight.com