POPULATION TRANSIENCE AS A DRIVER OF HOUSEHOLD RETAIL COSTS

Report for Thames Water

March 2018

1. Executive summary

This report for Thames Water (Thames) considers the extent to which population transience (transience) – the propensity of people to migrate – drives water companies' household (HH) retail costs.

We find evidence that transience, when appropriately defined, is a robust cost driver in some, but not all, econometric cost models – particularly those pertaining to bad debt-related costs.[1] As such, **we recommend that transience should be included within the pool of explanatory variables within any model suite used to assess company HH retail costs.**

We further find that, **should transience not be included in benchmarking models, firms facing 'low' transience might have allowed bad debt-related costs set at 4% <u>above</u> the efficient level**; whereas firms facing 'high' transience might have allowed bad debt-related costs set at 26% <u>below</u> the efficient level.

In addition, because many companies experience similar transience levels, with a smaller number of companies being 'outliers', it may be that: (a) certain econometric specifications fail to identify an 'across industry' cost relationship; and / or (b) even where econometric approaches do identify a relationship, these do not fully capture the impact of transience on some companies' costs. Consequently, in order **to ensure that efficient costs are allowed for, in addition to its inclusion in cost assessment models, it might also be necessary and appropriate for affected companies to make cost adjustment claims** relating to this issue.[2]

2. Introduction and scope of work

The objective of this report is to test the strength of evidence that transience should be incorporated within the cost assessment framework for HH retail at PR19. As part of our wider work on the HH retail control, we developed a set of econometric cost models, one of which includes a definition of transience as a statistically significant cost driver. This report tests the robustness of this finding to alternative measures of transience; and examines the performance of these measures within our wider suite of econometric models. In addition, to demonstrate the implication of omitting a valid driver of efficient costs, we analyse how company allowed costs are impacted by the exclusion / inclusion of transience.

The remainder of this report is structured as follows:

- Thames, who commissioned our work, provide a foreword, giving a company perspective on the issue of transience.

[1] *Where, for clarity, debt-related costs refer to the combination of doubtful debt and debt management.*
[2] *Subject to considering the criteria for cost adjustment claims, set out by Ofwat in the PR19 methodology, including adopting a 'symmetrical' approach.*

- We then set out our own economics perspective on how transience can drive HH retail costs.
- Next, we provide an analysis of how transience varies across companies, and the implications of this.
- We then set out the results of our econometric modelling, in which transience is found to be a statistically valid cost driver. This includes providing evidence of the robustness of our findings (diagnostic testing) and exploring whether the finding is consistent under various alternative model specifications.
- Finally, we summarise our conclusions.

3. Foreword from Thames Water

In its Final Methodology, Ofwat confirmed its intention to use econometric benchmarking to set companies' allowed revenues for the household retail controls. We welcome this approach.

Transience, in our experience, materially raises our customer service and bad debt costs. It increases our customer service costs, as greater transience implies that we need to accommodate more new customers. This involves, amongst other things, sending welcome packages and dealing with the queries of new customers. It raises our bad debt costs, as the risk of customers not paying their bills (and thus the potential scale of arrears) is greater for customers that move out of their properties (relative to customers staying in the same property). Moreover, recovering our revenues is more challenging and costly for customers that are on the move.

Analysis presented in this report establishes that transience (measured as the proportion of population migrating into/from local authorities) varies markedly across companies, and is particularly high for Thames Water. Figure 1 shows that transience in our area is 59% above the national average, and 30% greater than that of any other company. This shows that 'our' transience is a degree different. Also, to the extent that transience raises our costs, it points to the need to account for it in setting our allowed revenues (most naturally by including transience as a driver in the cost models).

With our experience of transience raising our costs and transience being so much greater in our area in mind, we commissioned Economic Insight to assess the impact of transience on companies' costs, using the household retail efficiency models it had developed. Economic Insight identifies transience as a statistically significant driver in one bad debt model across a range of transience measures. It also finds transience to have the expected sign, yet sometimes without being statistically significant, in its other bad debt models. These findings accord with our experience of transience materially raising the costs we incur in serving our household customers.

The best econometric models will be those that include all the relevant explanatory variables. Deciding which variables to include is not an exact science, but it is appropriate to consider all drivers that may materially affect companies' costs (when controlling for other factors) and vary markedly across companies. Based on this criterion, we recommend that Ofwat considers transience (measured as the propensity of people to migrate) as part of the mix of drivers it has regard to in developing its household retail models.

We trust readers will value this report's contribution to the development of robust household retail models, and we thank Economic Insight for its work.

Colm Gibson, Head of Regulation, Thames Water

March 2018

4. Transience as a driver of HH retail costs – our perspective

Transience is the propensity of people to migrate; and there are strong 'in principle' reasons to expect this to drive HH retail costs. We distinguish between both the *direction* of population movements, and their *geography*. With respect to direction, the key categories include: (i) inflows; (ii) outflows; and (iii) total flows (the sum of inflows and outflows). With respect to geography, it is useful to consider the following three cases.

- **Within-company transience**, refers to situations in which water companies' customers move properties, but remain *within the same company supply area.*
- **Within-UK transience** denotes situations in which customers move into, or out of, a company supply area – to / from *another location within the UK.*
- **International transience** refers to situations in which customers move into, or out of, a company supply area – to / from a location *outside of the UK*.

There are several distinct ways in which these different types of transience *would be expected to affect companies' HH retail costs* (in a manner that is beyond efficient management control). These require careful consideration – and so the remainder of this section discusses the various ways in which costs could be impacted, addressing: non-bad debt; then bad debt related, HH retail costs in turn.

4.1 How transience could affect non-bad debt related HH retail costs

With respect to **non-bad debt related costs**, the main impacts are on **account management** and **metering** costs – as follows:

- As the propensity of customers to move around increases, companies incur higher account management costs, associated with opening, closing or modifying customer accounts. In principle, this cost impact applies to <u>all</u> types of transience (inflows and outflows), although the extent of impact may vary by transience type.[3]
- Greater transience could be associated with a tendency for more customers to submit their own meter readings, as they move into new properties / leave existing properties, potentially reducing the number of meter reading trips that company operatives need to undertake. In principle, this applies most strongly to population inflows. In reality, this cost impact is expected to be limited.

[3] *For instance, setting up a new customer account (for someone who did not previously reside in a company's supply area, and so was not previously served by the company) may be more time consuming than simply changing a customer's address when they move within a company's supply area.*

4.2 How transience could affect bad debt related HH retail costs

TRANSIENCE CHANGES THE TRADE-OFF POINT BETWEEN DEBT MANAGEMENT AND DOUBTFUL DEBT – AND SO, FACED WITH INCREASED TRANSIENCE, COMPANIES UNAVOIDABLY FACE EITHER HIGHER DEBT MANAGEMENT OR DOUBTFUL DEBT COSTS.

With respect to **bad debt related costs**, it is important to keep in mind the inherent interdependencies between 'debt management' and 'doubtful debt' when considering the potential impact of transience. Specifically, as we explain below, transience very directly increases the costs of debt management activities *required to achieve a given level of doubtful debt.* In practice, companies have some discretion as to *whether* they 'choose' to incur those additional debt management costs, and will face a trade-off between: (i) investing more effort and resource (and therefore cost) in debt management; or (ii) having higher levels of doubtful debt. The issue is, however, that the impact of transience changes the trade-off point – and so it is unavoidable that, in the face of increased transience, companies will incur either higher debt management, or doubtful debt costs.

It is further important to think through the various ways in which transience can have these impacts – and how such impacts might vary across the various measures of transience we set out previously:

- **The more customers 'move around' the more difficult, time consuming, and costly, debt management activities are likely to be.** For example, if a customer in arrears moves location, companies are likely to incur additional costs both in tracing that customer and in recovering any monies owed. The extent of this impact is likely to vary by type of transience. For example, **the impact may be greater with respect to 'outflows'** – because the act of a customer, already in arrears or default, leaving by definition directly drives these costs. However, it might also be related to inflows for two reasons: (i) firstly, when thinking about 'within area' transience, an inflow should be 'matched' to an outflow – and therefore records the same event[4]; and (ii) secondly, if there is an association between overall migration and customers' propensity to get into arrears (say, because transience varies by socioeconomic group), then clearly, higher inflows could increase company debt management costs. In addition, the extent of impact could also vary by the 'geographic' definition of transience.[5]

- As noted above, to the extent that transience results in debt management becoming 'more difficult', the trade-off point with doubtful debt will change. Put simply, for a given level of debt management activity and cost, companies will face higher doubtful debt costs. In addition, when thinking specifically about doubtful debt, there may be a relationship between moving and falling into arrears. For example, if it takes time for customers that move into new properties to register with their supplier, this could lead them to 'build up' larger arrears and / or potentially tip them into default. This behaviour might be more prevalent amongst socioeconomic groups for whom arrears likelihood is already "high". Clearly, the impact of this would be further accentuated if there was also an underlying relationship between transience rates and socioeconomic groups.

---

[4] *Note, also, that in instances where, for example, an outflow might not be accurately recorded in the data, the 'inflow' measure ensures this is captured.*

[5] *For example, we might expect it to be easier to trace customers / recover monies relating to movements within a company supply area, relative to say, international migration.*

THE PRACTICAL IMPLICATION OF CONSIDERING HOW TRANSIENCE IMPACTS HH RETAIL COSTS IS THAT ONE SHOULD EXPLORE A RANGE OF MEASURES OF TRANSIENCE WHEN UNDERTAKING COST MODELLING.

4.3 Overview of expected impacts

Drawing on the above discussion, the following table summarises the expected impacts of transience on HH retail costs. Here, "double arrows" denote *larger* expected impacts, with "single arrows" denoting *smaller* expected impacts.

Table 1: Expected impact of transience on company HH retail costs

Geography	Direction	Account management	Metering	Debt management & Doubtful debt
Within-company	Inflow	↑↑	↓	↑
	Outflow	↑↑		↑↑
Within-UK	Inflow	↑↑	↓	↑
	Outflow	↑↑		↑↑
Inter-national	Inflow	↑↑	↓	↑
	Outflow	↑↑		↑↑

Source: Economic Insight

The main practical implications of the above for cost modelling are that: (i) one should consider a range of transience measures within an econometric setting; and (ii) one should ensure that whichever measure is used is sufficiently 'broad', so as to appropriately capture the various effects set out here.

This is because: (a) a careful consideration of how transience affects costs would indicate that multiple transience measures may be relevant; and (b) one might naturally expect alternative transience measures to be correlated (and so, even if there were intuitive reasons to favour one measure over another, from a statistical perspective these may not matter).

5. Examining transience across company supply areas

Data on transience in the UK is available for local authority areas from the Office for National Statistics (ONS). These distinguish between inflows and outflows; and between *internal flows*, which are population movements between UK local authorities, and *international flows*, to and from locations outside the UK. Data are not available on movements *within* UK local authorities. This generates nine transience measures, as set out in the table below.

Table 2: Transience measures

Variable		Description
A	Internal inflows	Inflows from other UK local authorities
B	Internal outflows	Outflows to other UK local authorities
C	Total internal transience	A + B
D	International inflows	Inflows from locations outside the UK
E	International outflows	Outflows to locations outside the UK
F	Total international transience	D + E
G	Overall inflows	A + D
H	Overall outflows	B + E
I	Overall transience	G + H

Source: Economic Insight

We used the above ONS data to calculate transience (for differing measures) by water company supply area. To do this, we undertook the following steps:

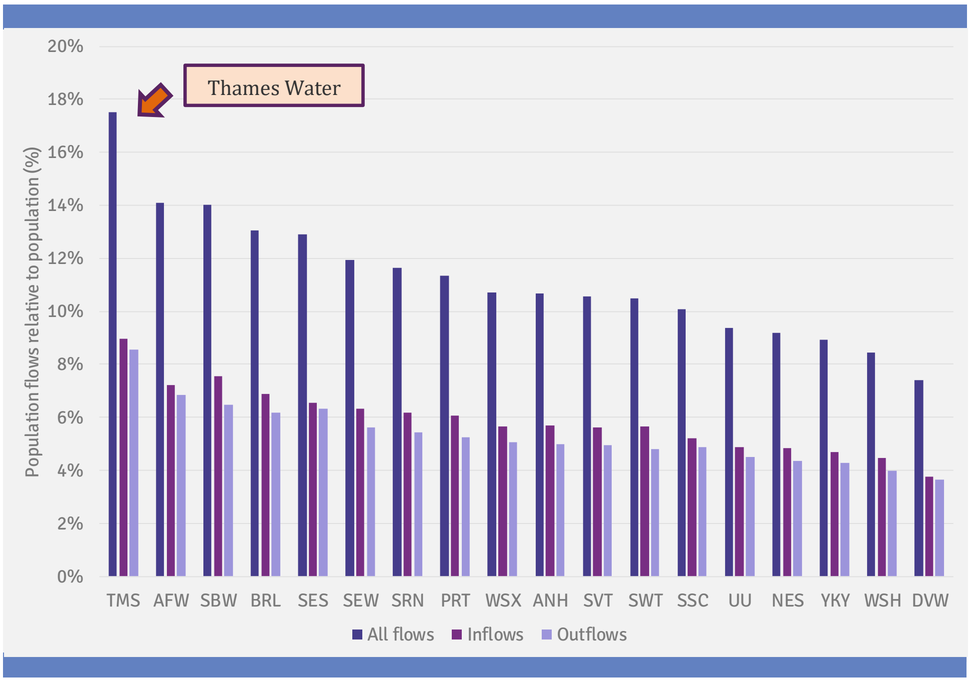
- Calculated the percentage geographical overlap between water supply areas and local authority geographical areas.
- Disaggregated local authority-level population flow data across companies, on the basis of these percentage overlaps, and summed these to generate company-level estimates of population movements.
- Divided the company-level population movements by supply area population, to control for scale.

Further details regarding our finalised measures of transience, including summary statistics for the raw data, are set out in Annex B.

DATA SHOWS THAT THERE IS A 'CLUSTERING' OF SIMILAR TRANSIENCE LEVELS FOR MOST COMPANIES, WITH 'OUTLIER' COMPANIES HAVING HIGHER TRANSIENCE.

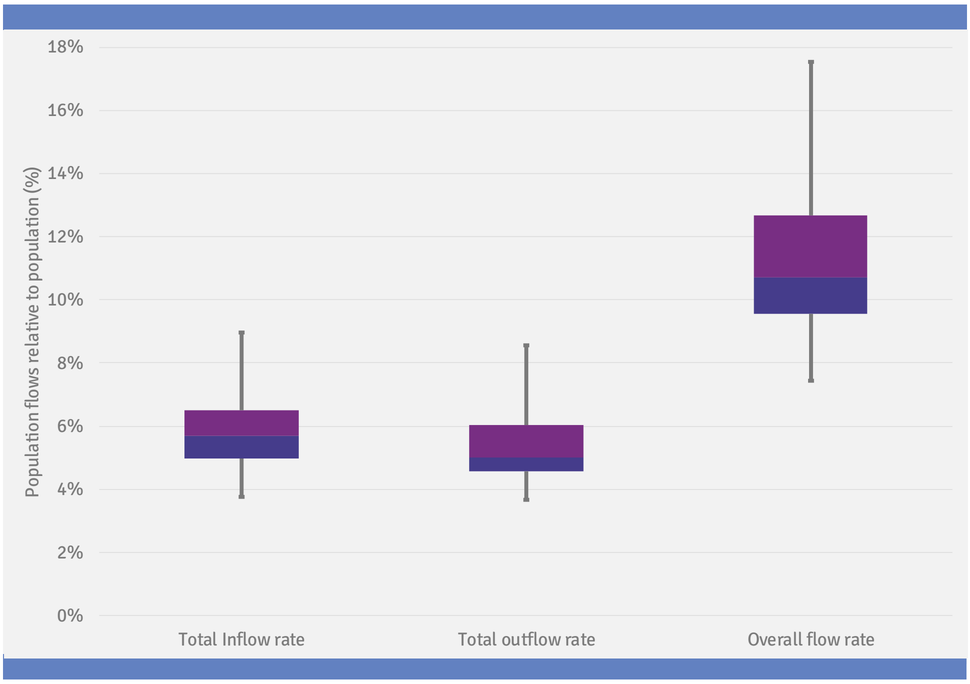
The following figure shows overall transience measures (inflows, outflows and total) across companies. There appears to be a high degree of 'clustering' acround the middle of the distribution, with a small number of companies having comparatively higher levels of transience. This can be seen in more detail in the accompanying box plots. These show narrow interquartile ranges (indicating that many companies have transience that is close to the median) but with much large absolute ranges (reflective of the fact that a small number of companies are outliers, with high transience).

Figure 1: Overall rates of transience



Source: Economic Insight

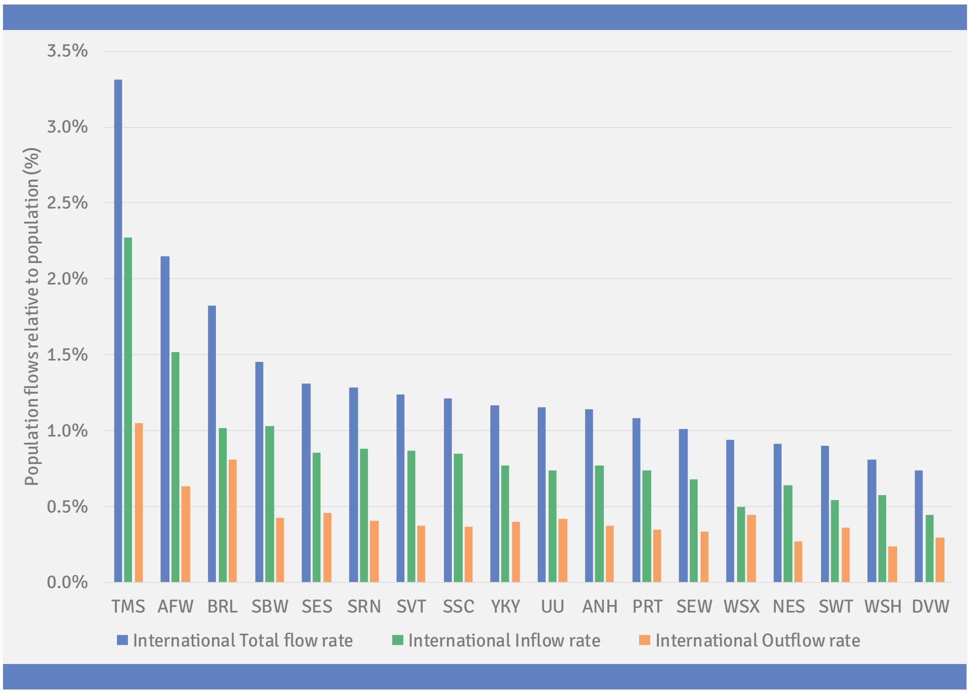
Figure 2: Box plots for overall rates of transience



Source: Economic Insight

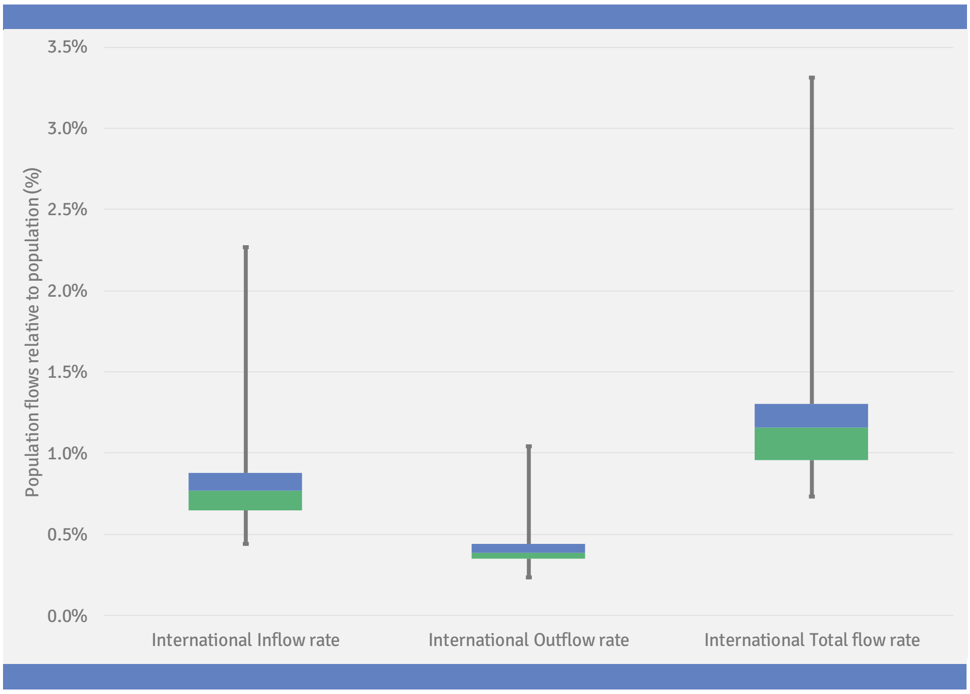
When considering **international transience** specifically, the contrast between the 'clustered' observations in the middle of the distribution, alongside a small number of companies with very high transience, is even more pronounced. We show this in the two figures below.

Figure 3: Rates of international transience



Source: Economic Insight

Figure 4: Box plots for international transience



Source: Economic Insight

5.1 Implications of ‘clustering’ for how transience could be reflected in PR19 cost allowances

THE INCLUSION OF TRANSIENCE IN COST BENCHMARKING MODELS ALONE MAY NOT BE SUFFICIENT TO ENSURE THAT TRANSIENCE’S IMPACT ON EFFICIENT COSTS IS ADEQUATELY CAPTURED.

As we set out in the following subsections, ultimately, we find that transience is a robust driver of HH retail costs within an econometric setting. This, combined with the fact that transience is outside of efficient management control, points to it being included within any suite of cost assessment econometric models used to set industry allowed costs at PR19.

Notwithstanding the above, it is important to also consider the implications of the apparent ‘clustering’ of companies with regard to transience, as highlighted in Figures 1 to 4 above. In particular, the data shows that, while there is variation in transience across companies, this is further characterised by: (i) a large number of companies having relatively similar transience levels; and (ii) a small number of companies being outliers (with much higher transience levels).

The practical implications of this are as follows:

- There may be certain specifications of econometric model that do not identify an ‘across industry’ transience effect.
- Even in specifications of econometric model that do identify an ‘across industry’ transience effect, the impact of this (as measured by any such model) may be understated, due to the clustering.
- Consequently, to fully ensure that the impact of transience on companies’ ‘efficient’ costs is included within the PR19 cost allowances, it might be necessary and appropriate for companies to make cost adjustment claims to Ofwat, addressing this issue.

6. Evidence on the impact of transience on HH retail costs from regressions

To inform an assessment of HH retail cost efficiency more broadly, we developed a suite of efficiency benchmarking models. As we explain below, these models were not developed for the purpose of examining ‘transience’ as a cost driver. Rather, they were arrived at through an objective general to specific approach, which was subject to academic peer review. Within our final suite of models, we found that the **total rate of internal transience** (i.e. the sum of inflows and outflows) was a statistically valid cost driver, within one of our bad debt related operating cost models.

Within the scope of this report, we have used our existing models as a starting point to further examine the validity of transience as a driver. This has included:

- Firstly, undertaking additional testing of the robustness of the inclusion of the total internal transience variable within the existing suite of models.
- Secondly, testing whether alternative measures perform better, or are more relevant, for other cost models in the suite.

6.1 Evidence from econometric benchmarking models

For context, we briefly summarise our approach to developing the suite of econometric cost models.

- We began with a first principles consideration of the drivers of HH retail costs. This included transience, alongside other drivers such as scale (customer numbers), scope (dual versus single service customers), meter penetration, traffic speed and wholesale bill amount.
- We matched these cost drivers to available data, including the transience data that we describe above, alongside information from the company data share, other ONS data on socioeconomic factors, and Department for Transport traffic data.
- We used a general to specific modelling approach to estimate a suite of 16 models. In doing this, we balanced intuition with statistical significance by retaining variables that were statistically significant at levels approaching 10%. This is primarily driven by a concern that a stricter approach would risk omitted variable bias, as scale is such a dominant driver of HH retail costs.
- Half of the models addressed scale and scope factors through the inclusion of separate variables for the numbers of single and dual service customers (model set A); the other half did so using separate variables for the total number of customers and the number of single service customers (model set B).
- We estimated eight total HH retail cost models, half of which included additional variables that were not statistically significant (evaluated at up to 15%) but which were correctly signed (we label this our *‘alternative approach’*), alongside four models each for bad debt and non-bad debt costs. We estimated pooled versions

of each model using ordinary least squares (OLS), and random effects versions, using generalised least squares (GLS).[6]

The final suite of models is summarised in the table overleaf.

Table 3: Suite of econometric benchmarking models

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

Source: Economic Insight

[6] *Our dataset has a panel structure, including repeated observations of the same companies over several years. This creates a potential statistical issue, in that unadjusted OLS standard errors may be under-estimated. Several options are available to address this, including estimating random effects models, or using clustered standard errors. However, these approaches also have drawbacks – particularly in the context of small sample sizes. As such, there are risks associated with relying exclusively on pooled OLS models (which do not address clustering) on the one hand, and random effects models, or models with clustered standard errors (which do address clustering) on the other. Our approach, therefore, is to include both unadjusted OLS standard errors and random effects models within our suite – so as to balance these considerations. This is consistent with Ofwat's approach at PR14.*

THE STATISTICAL SIGNIFICANCE OF POPULATION TRANSIENCE IN MODEL A2 IMPLIES THAT IT IS A VALID COST DRIVER, WITH THE POTENTIAL TO MATERIALLY IMPACT FIRMS' MODELLED COSTS.

Following general to specific modelling, the total rate of internal transience (i.e. both inflows and outflows) was included within model A2 (pooled OLS, bad debt related costs). The following tables summarise: (i) the form of the model (for further details of variables and sources, please see annex C); (ii) model results; and (iii) model diagnostics, *in a format consistent with that outlined by Ofwat in relation to its forthcoming consultation on cost assessment models.*

Table 4: Model A2 – summary of model form

Model form
Description of econometric model formula: $\ln(\textit{bad debt related costs}_{it}) = \beta_0 + \beta_1 \ln(\textit{single service customers}_{it}) + \beta_2 \ln(\textit{dual service customers}_{it}) + \beta_3 \textit{ IMD income}_{it} + \beta_4 \ln(\textit{average wholesale bill}_{it}) + \beta_5 \textit{ total internal migration}_{it} + \varepsilon_{it}$
Description of dependent variable: Bad debt related operating costs (doubtful debt and debt management costs).

Source: Economic Insight,

The model results are shown in the table below. It suggests that a 1% increase in the rate of total internal migration is, on average, associated with a 0.09% increase in bad debt related HH retail costs. The coefficient is statistically significant, implying that, in this model, transience is a valid cost driver, with the potential to have a material impact on firms' implied costs.

Table 5: Model A2 – model results

Variables	Model A2 Bad debt costs (ln)	
Single service customers (ln)	Coefficient	0.535***
	P-value	(0.000)
Dual service customers (ln)	Coefficient	0.121***
	P-value	(0.000)
IMD income (%)	Coefficient	0.189***
	P-value	(0.000)
Wholesale bill (ln)	Coefficient	1.744***
	P-value	(0.000)
Total internal migration (%)	Coefficient	0.0909***
	P-value	(0.001)
Constant	Coefficient	-14.37***
	P-value	(0.000)

*Source: Economic Insight, ***p<0.01, ** p<0.05, * p<0.1; p-values use robust standard errors*

6.2 Diagnostic tests – and model information

We applied a range of diagnostic tests to this model. The table below summarises results consistent with those set out in Ofwat's template for the forthcoming cost assessment consultation.[7]

Table 6: Diagnostic tests applied to model A2

Test	Statistic	Implication
R^2 adj.	0.933	Model explains approximately 93% of the variation in the dependent variable.
RESET test	0.0004	None – see below.
Variance Inflation Factor (VIF)	3.55 (mean) 6.78 (max)	Low VIF (below 10) does not indicate multicollinearity issues.
Method (e.g. OLS or RE)	Pooled OLS	
N (sample size) Companies Years	89 18 5	

Source: Economic Insight

Relating to the above, if model A2 is re-run without the transience variable included, the R2 value falls to 0.927 (i.e. the overall explanatory power of the model is reduced, where transience is not captured).

[7] *We suggest placing limited weight on the Ramsey Regression Equation Specification Error Test (RESET) in the context of efficiency modelling. The test is likely to have difficulty in discriminating between genuine omitted variable or functional form problems, and cases in which cost model residuals are 'correctly' measuring inefficiency. For example, adding a variable that was correlated with inefficiency to a model that was otherwise correctly specified would lead to the model performing 'better' on RESET, even though this would lead the model to understate inefficiency.*

6.3 Comments on the overall performance of model A2

We consider A2 to be a credible model of company bad debt related operating costs. It includes a range of intuitively sensible drivers of doubtful debt and debt management costs, all of which are strongly statistically significant. Diagnostics suggest that it explains over 90% of the variation in the dependent variable, and there are no indications of problems relating to multicollinearity of the variables. As we described above, our suite of econometric models included both pooled and random effects models, alongside an alternative approach to modelling customer numbers. In both respects we think that it is useful to include model A2 as part of the model suite, alongside these alternative approaches.

Specifically, with respect to incorporating the panel structure of the data within the econometric models, we regard pooled models (such as A2) and random effects models as complementary. While random effects models have the advantage of distinguishing between statistical noise and other error components, the resulting inefficiency estimates are invariant across time. In contrast, pooled models do not make a distinction between noise and other error components, but are able to provide time-varying inefficiency estimates.

Turning to the approach to modelling customer numbers, the inclusion of separate dual and single service customer variables (as in model A2) provides a very flexible specification. The alternative of separate total and single service customer variables is more restrictive, though in practice it avoids complications around the treatment of firms with no dual service customers. As such, we think that both approaches are valid and should be included within a suite of models.

6.4 Scale of impact

Our comparison of transience data across companies indicated that there is a wide range of transience levels, though most are concentrated towards the centre of the distribution. As such, the inclusion or exclusion of transience has the potential to have a material impact on firms' implied efficient costs.

To examine this further, we compared total predicted bad debt related costs in model A2 with a version of the model that excludes the transience variable. The table below shows total predicted bad debt related costs (per annum) from these models, for all firms with: levels of transience below the lower quartile; above upper quartile; and for Thames separately.

SHOULD TRANSIENCE NOT BE INCLUDED IN BENCHMARKING MODELS, FIRMS WITH HIGH TRANSIENCE MIGHT HAVE ALLOWED COSTS SET 26% BELOW THE EFFICIENT LEVEL.

Table 7: Change in predicted costs when transience is excluded from modelling

Firms	Total internal transience level (%)	Predicted costs excluding transience variable (£m)	Predicted costs including transience variable (£m)
Below lower quartile transience	<8.4%	£144.4	£138.7
Above upper quartile transience	>11.2%	£78.8	£105.9
Thames Water	14.2%	£68.8	£95.3

Source: Economic Insight

The implications of the above for HH retail cost assessment are significant. In particular, it indicates that, should transience not be included within benchmarking models:[8]

- Firms facing 'low' transience might have their allowed (debt related) costs set **around 4% higher than the 'efficient' level.**
- Firms that face 'high' transience might have their allowed (debt related) costs set **around 26% lower than the 'efficient' level.**
- For Thames specifically, the impact of excluding transience would result in its (debt related) costs being set at **around 28% below the 'efficient' level.**

Clearly, in practice the 'full' impact on allowed costs of including / omitting transience would depend on a number of other considerations – including how results were weighted across multiple models under triangulation.

[8] *The following figures reflect the 'omission' of transience from model A2.*

6.5 Testing alternative transience measures of transience

TRANSIENCE TENDS TO BE SIGNIFICANT WITHIN DEBT RELATED COST MODELS.

We tested all nine measures of transience within the models described above. The tables overleaf summarise the results across the suite of models. We have used: (i) **green dots** to denote cases in which the transience variable was statistically significant at 5%; (ii) **orange dots** to show cases in which it was statistically significant at between 5% and 10%; and (iii) **purple dots** to show cases in which it was statistically significant at between 10% and 15%.

- It is within the debt related cost models that transience is most often statistically significant. This applies most strongly in the case of pooled OLS models (A2 and B2), although there are some indications of statistical significance across the set A random effects debt model (A6).

- Looking more closely at models A2 and B2, while all of the transience measures are statistically significant at 5% for A2, the measures that include *international transience* are statistically significant at this level for model B2. This would seem to accord with our previous discussion of the intuition relating to how transience can impact HH retail costs – which indicated that: (i) all transience measures may affect HH retail costs, to differing degrees; and (ii) that, even if there were intuitive reasons to 'prefer' a particular measure (i.e. inflows or outflows) in certain specific contexts, from a statistical perspective this is unlikely to matter.

- Relating to the above, expressing transience as inflows or outflows does not appear to have a material impact on statistical significance, except in the case of A3. This is likely to reflect the fact that, in practice, inflows and outflows are highly correlated.

- In some cases, the non-bad debt related cost models (models A3, A6 and B3) are also statistically significant, though this is less consistent. The coefficients for models A3 and B3 (non-bad debt related costs, pooled OLS) are negative. This may accord with the idea that transience could be associated with reduced metering costs, or it could reflect correlation with other cost drivers (for instance with how 'urban' a supply area is).

- The lack of significance of transience in the total cost models (apart from A1) may reflect the tendency of scale variables to dominate all other explanatory variables, particularly at very aggregated levels of cost.

- In general, the transience variables are more frequently statistically significant in the pooled OLS models, and in model set A.

Table 8: Alternative transience measures: models A1 to A8

Variable	Model A1	Model A2	Model A3	Model A4	Model A5	Model A6	Model A7	Model A8
Internal inflow (%)	🟣	🟢	🟢			🟣		
Internal outflow (%)	🟣	🟢	🟣			🟣		
Internal overall flow (%)	🟣	🟢	🟢			🟣		
International inflow (%)		🟢						
International outflow (%)		🟢				🟣		
International total flow (%)		🟢						
Overall inflow (%)		🟢	🟢					
Overall outflow (%)		🟢	🟣			🟠		
Overall total flow (%)		🟢	🟢			🟣		

Source: Economic Insight

Table 9: Alternative transience measures: models B1 to B8

Variable	Model B1	Model B2	Model B3	Model B4	Model B5	Model B6	Model B7	Model B8
Internal inflow (%)			🟢					
Internal outflow (%)		🟣						
Internal overall flow (%)		🟣						
International inflow (%)		🟢						
International outflow (%)		🟢						
International total flow (%)		🟢						
Overall inflow (%)		🟠	🟣					
Overall outflow (%)		🟠						
Overall total flow (%)		🟠						

Source: Economic Insight

‘if Ofwat develops a suite of econometric models for HH retail cost assessment, we would recommend that transience should be included within the broader mix of explanatory variables.’

7. Conclusions

As we have set out, there are strong intuitive reasons to suggest that transience can impact water companies’ HH retail operating costs in a manner that is outside of efficient management control. The evidence we have examined here shows that: (i) the inclusion of transience within our econometric cost modelling is robust; and that, more broadly (ii) it is possible to identify a range of transience measures as valid cost drivers, within alternative forms of econometric models.

Following from the above, if Ofwat develops a suite of econometric models for HH retail cost assessment, we would recommend that transience should be included within the broader mix of explanatory variables. Importantly, we further show that the impact of excluding transience on company cost allowances can be material. Consequently, were it to be excluded:

- costs for companies with ‘low’ transience might be set above the efficient level (resulting in their customers ‘overpaying’); and
- costs for companies with ‘high’ transience, might be set below the efficient level (which might lead to ‘inefficient’ cost cutting, which in turn might be detrimental to the quality of service received by customers).

Our analysis also shows, however, that the level of variation in transience across companies is characterised by (i) a ‘clustering’ of companies with similar transience; and (ii) a smaller number of outliers, with very high transience. As such, some econometric specifications may fail to identify ‘across industry’ cost relationships. In addition, even in circumstances in which econometric models do identify a transience / cost relationship, the ‘clustering’ effect may mean that the full impact on company costs is understated. As such, it may further be necessary and appropriate for companies to consider cost adjustment claims in order to ensure that efficient costs are allowed for.

8. Annex A: Econometric models including alternative transience measures

The tables below provide full details of the econometric estimates of the alternative transience measures within the econometric cost models.

Table 10: Alternative transience measures: models A1 to A8

Variable	Statistic	Model A1	Model A2	Model A3	Model A4	Model A5	Model A6	Model A7	Model A8
Internal inflow (%)	Coefficient	0.0610	0.185***	-0.113***	0.0289	0.0212	0.196	-0.0615	-0.0340
	Standard error	(0.0419)	(0.0541)	(0.0420)	(0.0490)	(0.0840)	(0.128)	(0.0904)	(0.0914)
	P-value	0.149	0.000954	0.00871	0.557	0.801	0.126	0.496	0.710
Internal outflow (%)	Coefficient	0.0733	0.169***	-0.0701	0.0438	0.0699	0.174	-0.000206	0.0154
	Standard error	(0.0491)	(0.0505)	(0.0461)	(0.0583)	(0.0780)	(0.113)	(0.0951)	(0.0910)
	P-value	0.139	0.00121	0.132	0.455	0.370	0.124	0.998	0.866
Internal overall flow (%)	Coefficient	0.0348	0.0909***	-0.0488**	0.0188	0.0287	0.101	-0.0192	-0.00590
	Standard error	(0.0230)	(0.0264)	(0.0213)	(0.0269)	(0.0445)	(0.0628)	(0.0503)	(0.0516)
	P-value	0.134	0.000896	0.0245	0.487	0.519	0.109	0.704	0.909
Internat-ional inflow (%)	Coefficient	0.128	0.401***	0.0528	0.0828	0.0786	0.108	0.0242	-0.00691
	Standard error	(0.136)	(0.116)	(0.175)	(0.171)	(0.121)	(0.206)	(0.203)	(0.140)
	P-value	0.349	0.000822	0.764	0.630	0.517	0.600	0.905	0.961
Internat-ional outflow (%)	Coefficient	0.131	0.858***	-0.538	0.0668	0.238	0.582	0.281	0.183
	Standard error	(0.253)	(0.236)	(0.371)	(0.310)	(0.206)	(0.399)	(0.239)	(0.209)
	P-value	0.606	0.000477	0.151	0.830	0.247	0.144	0.240	0.382
Internat-ional total flow (%)	Coefficient	0.0778	0.291***	-0.0364	0.0480	0.105	0.160	0.155	0.0473
	Standard error	(0.0935)	(0.0818)	(0.125)	(0.117)	(0.0976)	(0.160)	(0.167)	(0.111)
	P-value	0.408	0.000634	0.772	0.683	0.281	0.320	0.353	0.671
Overall inflow (%)	Coefficient	0.0469	0.138***	-0.0752**	0.0254	0.0268	0.109	-0.0359	-0.0188
	Standard error	(0.0351)	(0.0382)	(0.0331)	(0.0422)	(0.0566)	(0.0869)	(0.0723)	(0.0654)
	P-value	0.184	0.000531	0.0258	0.550	0.636	0.210	0.619	0.773
Overall outflow (%)	Coefficient	0.0581	0.150***	-0.0707	0.0351	0.0836	0.167*	0.0374	0.0402
	Standard error	(0.0439)	(0.0433)	(0.0429)	(0.0531)	(0.0696)	(0.0982)	(0.0879)	(0.0816)
	P-value	0.189	0.000871	0.103	0.511	0.229	0.0884	0.671	0.622
Overall total flow (%)	Coefficient	0.0264	0.0726***	-0.0375**	0.0151	0.0280	0.0718	-0.00372	0.00248
	Standard error	(0.0196)	(0.0204)	(0.0187)	(0.0237)	(0.0331)	(0.0476)	(0.0424)	(0.0394)
	P-value	0.182	0.000620	0.0478	0.525	0.398	0.131	0.930	0.950

Source: Economic Insight

Table 11: Alternative transience measures: models B1 to B8

Variable	Statistic	Model B1	Model B2	Model B3	Model B4	Model B5	Model B6	Model B7	Model B8
Internal inflow (%)	Coefficient	0.0131	0.0625	-0.0501**	0.00660	-0.0250	0.0395	-0.0525	-0.0357
	Standard error	(0.0232)	(0.0451)	(0.0237)	(0.0213)	(0.0461)	(0.0984)	(0.0569)	(0.0480)
	P-value	0.573	0.169	0.0374	0.758	0.587	0.688	0.356	0.457
Internal outflow (%)	Coefficient	0.0163	0.0552	-0.00512	0.00779	-0.00461	0.00540	-0.000643	-0.0153
	Standard error	(0.0212)	(0.0370)	(0.0274)	(0.0196)	(0.0453)	(0.0879)	(0.0647)	(0.0475)
	P-value	0.443	0.140	0.852	0.693	0.919	0.951	0.992	0.747
Internal overall flow (%)	Coefficient	0.00754	0.0302	-0.0156	0.00368	-0.00792	0.0112	-0.0162	-0.0139
	Standard error	(0.0112)	(0.0205)	(0.0128)	(0.0103)	(0.0238)	(0.0483)	(0.0318)	(0.0250)
	P-value	0.502	0.145	0.226	0.723	0.739	0.817	0.610	0.579
Internat-ional inflow (%)	Coefficient	0.0392	0.176**	0.0559	0.00877	-0.00744	0.0476	0.0758	-0.0321
	Standard error	(0.0472)	(0.0693)	(0.0617)	(0.0416)	(0.0858)	(0.164)	(0.150)	(0.0909)
	P-value	0.409	0.0128	0.368	0.834	0.931	0.771	0.612	0.724
Internat-ional outflow (%)	Coefficient	-0.0709	0.400**	-0.266	-0.0520	0.0404	0.160	0.125	0.0379
	Standard error	(0.100)	(0.173)	(0.189)	(0.0893)	(0.168)	(0.342)	(0.209)	(0.170)
	P-value	0.481	0.0234	0.162	0.562	0.810	0.639	0.552	0.824
Internat-ional total flow (%)	Coefficient	0.0106	0.131**	-0.00244	-0.00215	0.00208	0.0496	0.0861	-0.0130
	Standard error	(0.0332)	(0.0520)	(0.0459)	(0.0290)	(0.0681)	(0.127)	(0.116)	(0.0715)
	P-value	0.750	0.0137	0.958	0.941	0.976	0.695	0.458	0.855
Overall inflow (%)	Coefficient	0.0111	0.0520*	-0.0302	0.00437	-0.0145	0.0261	-0.0274	-0.0246
	Standard error	(0.0160)	(0.0281)	(0.0192)	(0.0144)	(0.0338)	(0.0670)	(0.0471)	(0.0357)
	P-value	0.489	0.0676	0.120	0.763	0.668	0.697	0.560	0.491
Overall outflow (%)	Coefficient	0.00924	0.0530*	-0.0114	0.00371	-0.00123	0.0120	0.00946	-0.00959
	Standard error	(0.0180)	(0.0313)	(0.0240)	(0.0164)	(0.0402)	(0.0772)	(0.0584)	(0.0421)
	P-value	0.609	0.0937	0.637	0.821	0.976	0.876	0.871	0.820
Overall total flow (%)	Coefficient	0.00523	0.0266*	-0.0115	0.00207	-0.00475	0.0106	-0.00693	-0.00969
	Standard error	(0.00851)	(0.0148)	(0.0107)	(0.00769)	(0.0188)	(0.0368)	(0.0269)	(0.0198)
	P-value	0.540	0.0761	0.287	0.788	0.800	0.774	0.797	0.625

Source: Economic Insight

9. Annex B: Further details of our measures of transience

As explained in the main body of this report, we used ONS data on transience to derive transience measures by company supply area, based on percentage geographical overlaps between local authorities and water supply areas. Summary statistics for these data are shown in the table below. We note that the internal transience measures do not correspond precisely with the geographical distinctions set out in section 5. These measures will pick up within-company movements that occur between two local authorities that are both within a company supply area but will not pick up within-company movements that occur inside the same local authority area. Similarly, within-UK movements will only be captured to the extent that companies' supply areas cover different local authorities.

Table 12: Summary statistics for transience variables across companies

Variable	Mean	Standard deviation	Minimum	Maximum
Internal inflows (%)	5.04%	0.94%	3.21%	6.96%
Internal outflows (%)	4.91%	1.00%	3.28%	7.58%
Total internal transience (%)	9.95%	1.91%	6.49%	14.52%
International inflows (%)	0.82%	0.40%	0.38%	2.34%
International outflows (%)	0.44%	0.18%	0.20%	1.06%
Total international transience (%)	1.26%	0.57%	0.62%	3.33%
Overall inflows (%)	5.86%	1.25%	3.63%	9.26%
Overall outflows (%)	5.35%	1.15%	3.48%	8.61%
Overall transience (%)	11.21%	2.38%	7.11%	17.84%

Source: Economic Insight

10. Annex C: further details of variable definitions and sources relating to Model A2

The table below provides further information regarding the precise measures used, and sources for, our dependent and explanatory variables used in Model A2.

Table 13: Further details of variables in Model A2

Variable	Measure used	Source
Debt related costs (dependent)	Sum of doubtful debt and debt management costs.	Company data share.
Single service customers	Number of water-only and wastewater-only customers.	Company data share.
Dual service customers	Number of water and wastewater customers.	Company data share.
IMD income	The income domain score from the IMD.	IMD England (2015); IMD Wales (2014).
Average wholesale bill	Average wholesale bill.	Company data share.
Total internal migration	Sum of internal migration inflows and internal migration outflows, relative to population.	Office for National Statistics.

Source: Economic Insight

Economic Insight Limited

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

Economic Insight Ltd is registered in England No. 7608279.

Whilst every effort has been made to ensure the accuracy of the material and analysis contained in this document, the Company accepts no liability for any action taken on the basis of its contents. Economic Insight is not licensed in the conduct of investment business as defined in the Financial Services and Markets Act 2000.

Any individual or firm considering a specific investment should consult their own broker or other investment adviser. The Company accepts no liability for any specific investment decision, which must be at the investor's own risk.

© Economic Insight, 2018. All rights reserved. Other than the quotation of short passages for the purposes of criticism or review, no part of this document may be used or reproduced without express permission.

Economic Insight