Deloitte.

# Customer Service Incentives: statistical exploration of the water industry SIM

## Final Report for Thames Water Utilities Limited

**27 May 2016**

# Important Notice

This final report (the "Final Report") has been prepared by Deloitte LLP ("Deloitte") for Thames Water Utilities Limited ("TWUL") in accordance with the contract with them dated 15 December 2015 ("the Contract") and on the basis of the scope and limitations set out below.

The Final Report has been prepared solely for the purposes of testing the statistical validity of Service Incentive Mechanism's calculation, as set out in the Contract. It should not be used for any other purpose or in any other context, and Deloitte accepts no responsibility for its use in either regard.

The Final Report is provided exclusively for TWUL's use under the terms of the Contract, however it may be made available to Ofwat solely for the purpose of evaluating the statistical validity of Service Incentive Mechanism's calculation. No party other than TWUL, including Ofwat, is entitled to rely on the Final Report for any purpose whatsoever and Deloitte accepts no responsibility or liability or duty of care to any party other than TWUL in respect of the Final Report or any of its contents. If Ofwat chooses to rely on the Final Report, it does so at its own risk and without recourse to Deloitte.

As set out in the Contract, the scope of our work has been limited by the time, information and explanations made available to us. The information contained in the Final Report has been obtained from TWUL and third party sources that are clearly referenced in the appropriate sections of the Final Report. Deloitte has neither sought to corroborate this information nor to review its overall reasonableness. Further, any results from the analysis contained in the Final Report are reliant on the information available at the time of writing the Final Report and should not be relied upon in subsequent periods.



Any decision to invest, conduct business, enter or exit the markets considered in the Final Report should be made solely on independent advice and no information in the Final Report should be relied upon in any way by any third party. This Final Report and its contents do not constitute financial or other professional advice, and specific advice should be sought about your specific circumstances. In particular, the Final Report does not constitute a recommendation or endorsement by Deloitte to invest or participate in, exit, or otherwise use any of the markets or companies referred to in it. To the fullest extent possible, both Deloitte and TWUL disclaim any liability arising out of the use (or non-use) of the Final Report and its contents, including any action or decision taken as a result of such use (or non-use).

# Contents

# Executive Summary

Deloitte LLP ("Deloitte") has been commissioned by Thames Water Utilities Limited ("TWUL") to review the statistical methodology used in the calculation and application of Service Incentive Mechanism ("SIM") scores. The SIM has two components:

- a quantitative component, which tracks the number of complaints and unnecessary contacts received by each water company; and
- a qualitative component, which measures the customer satisfaction levels with the water company services. The customer satisfaction scores ("CSAT") scores are based on surveys of customers who contacted the water company.

The objective of this work was to use econometric analysis to assess the current design of the SIM, to explore the underlying statistical approach and to consider whether this has any implications for different water companies.

A combination of TWUL and publicly available information was used to develop the datasets used to test the various customer, contact and company specific attributes that may influence the customer satisfaction with the water companies measured through CSAT. Only the CSAT scores have been used in the econometric analysis.

## Econometric analysis of CSAT scores

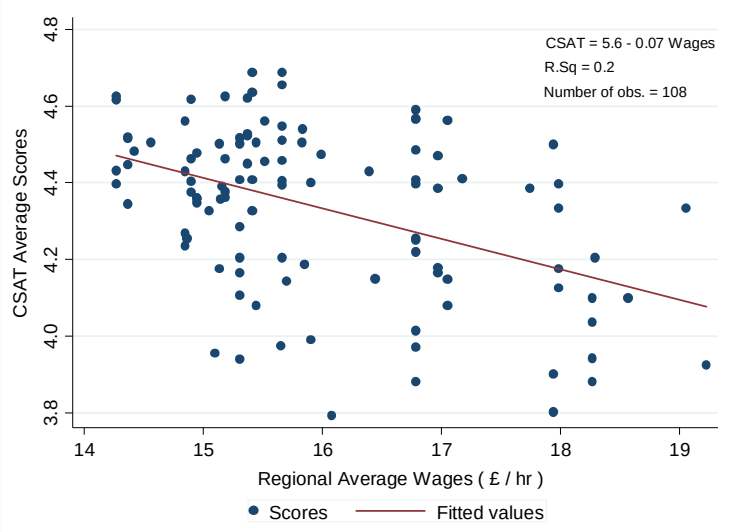
A series of regressions were used to test the soundness of the methodology used to sample customers across England and Wales companies. This analysis identified a number of customer, contact[1] and company characteristics that are associated with different CSAT scores in a statistically significant way. In particular:

- **Customer attributes.** Lower customer income or wages in the company's area are associated with higher CSAT scores. Many of the other socio-economic characteristics tested are less consistent in terms of statistical significance of direction of the impact on CSAT scores across model specifications.
- **Contact attributes.** Surveys that take place over a longer period of time are associated with higher CSAT scores, but this effect disappears when controlling for whether the contact has been resolved. Other contact attributes, such as whether it was a complaint, whether it was in writing and the reason for contacting the water company are also significant factors in explaining the CSAT scores.
- **Company attributes.** Water-only companies are associated with higher scores compared to Water and Sewerage companies, but other company characteristics, including geographical location and average company expenditure per customer tend not to be statistically significant.

Simple correlations illustrating the relationships above are provided below, although the findings themselves are based on the econometric analysis presented in the full report.

[1] Samples for the CSAT surveys are based on the list of customers who contacted the company within the month preceding the survey. Customers are not randomly selected from the wider customer population of each water company.

**Figure 1: Simple correlations among key SIM variables: average CSAT scores across companies and regional wages**




*Source: Deloitte analysis*

*Notes:*

*1. The data reflects annual average CSAT scores from 2010 Q1 to 2015 Q4.*

*2. In interpreting the diagram, if there was no statistical relationship between CSAT scores and the regional wages, the $R^2$ metric reported in the figure would be expected to be zero*

On the basis of the results of the econometric analysis, the SIM scores would be expected to systematically vary across companies independently of water companies' quality of service. As a result, companies may be rewarded and/or penalised under the existing SIM methodology not just for the quality of service provided to their customers, but also for factors that are largely beyond their control, such as the socio-economic characteristics of the population in their appointed area.

In particular, companies whose customers live in areas with lower average income (or wages) tended to be associated, other things being equal, with a higher CSAT (and hence SIM) score. Under the current SIM methodology, Ofwat would be rewarding such companies for the higher customer satisfaction. As a result, the water companies' revenues (and potentially customer water bills) might increase for those customers who live in areas with lower average income.

## Sample size

Based on the standard methodology used in household surveys, a conventional size of the quarterly CSAT sample size, specifically for TWUL in a single quarter, would include 380 households. This is greater than the current sample size of around 200 consumers used in the quarterly CSAT surveys.

The size of the sample selected is one of the most important features of the overall design of any customer satisfaction survey. The convention used in many household surveys[2] is to choose a sample size that depends on specific parameters, such as the margin of error in the underlying data and the desired level of confidence in the results. The sample size would also depend on the previous estimates of the target indicator (such as the customer satisfaction score), the anticipated rate of non-response to the survey and the share of the target population in the total population.

[2] UN Guidelines: Chapter 3: UN Department of Economic and Social Affairs – Statistics Division. Designing Household Survey Samples: Practical Guidelines, 2005

## CSAT methodology: raw and normalised scores

Building on the findings of an existing customer satisfaction research literature, it may not be appropriate to use raw scores from customer satisfaction surveys to directly compare water companies because of the underlying differences in customer expectations across the companies' respective customer pools.[3]

Applying a normalisation approach to customer satisfaction scores helps to address the inherent differences in customer expectations and characteristics across companies. If this approach was applied to CSAT surveys carried out in the past the relative ranking of water companies would have been different.

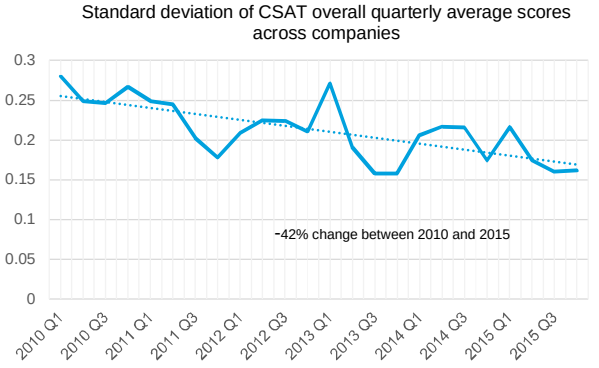
There are a number of different potential normalisation techniques that can be applied in relation to CSAT scores, although data availability puts a limit on the type of analysis that can be carried out. This report illustrates a simplified approach to provide the intuition for the normalisation methodology.

The results, based on the normalisation approach illustrated in section 3.2.3, indicated that companies that operate in areas with higher average incomes could have their normalised scores increase relative to those that operate in areas with lower average incomes. As a result of applying an alternative normalisation methodology, the ranking of the companies would be expected to change, with companies operating in areas with higher income tending to improve their relative position in the ranking.

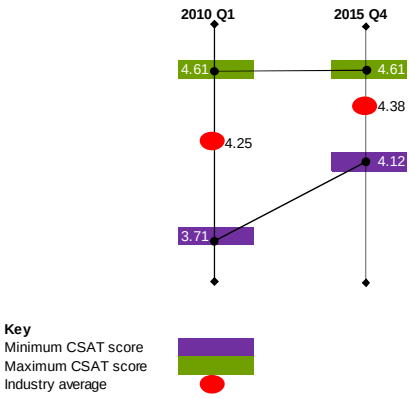
## Convergence of CSAT over time

The variability in CSAT survey scores across companies has been declining over time as shown in Figure 2. The decrease in the standard deviation of CSAT scores around the industry average implies a disproportionate impact of penalties for the water companies ranked below the industry average, as the distance between the lowest and highest ranked companies has declined over time.

**Figure 2: CSAT Standard deviation over time**




*Source: Overall weighted average scores 2010 Q1 - 2015 Q4, McCallum Layton League Tables*




[3] If consumers in different areas indeed have diverging expectations in terms of the quality of service from their water company, then a direct comparison of companies' scores fails to indicate how well companies perform in meeting their own customers' expectations.

# 1 Introduction and Background

## 1.1 Introduction and background

Since 2010, water and sewerage companies in England and Wales have been subject to an industry-wide incentive set by the water sector regulator Ofwat known as the Service Incentive Mechanism ("SIM"). The SIM was introduced as a way to provide comparable measures of customer satisfaction levels for the services rendered by the 18 water and wastewater companies across England and Wales and to provide an incentive for companies to improve customer satisfaction with those services.

The quarterly satisfaction surveys (CSAT) explore the customers' views on their recent experience with their water supplier. Each individual customer who contacted the water company is asked to rate her overall satisfaction with their water company on a scale from 1 to 5, where 1 means "very dissatisfied" and 5 means "very satisfied". The latest wave of customer surveys carried out in Q4 2015/16 has included several structural changes, such as: the exclusion of non-household contacts, the selection of samples from all contacts, regardless of whether their issue was resolved or not at the time of the survey and the satisfaction scores are weighted based on the water companies' respective services.

The outcome of these surveys is an overall customer satisfaction level for each water company, which is used by Ofwat to rank those companies and reward or penalise them based on their ranking relative to the average across all companies. Therefore, the soundness of the methodology underlying SIM is very important both for companies and the regulator. The SIM is composed of two elements: the CSAT scores based on customer surveys together with a score on unwanted contacts that each company receives. This report focuses primarily on the CSAT scores.

In this context, TWUL has commissioned Deloitte to review the statistical methodology used in the calculation and application of SIM scores.

## 1.2 Scope of work

Two main aspects of the survey methodology are considered: the soundness of the methodology used to sample customers across the companies and the appropriateness of using the satisfaction scores directly to compare and rank companies.

In support of this work, the following analysis was conducted:

- The formulation of statistical hypotheses to test the soundness of the CSAT survey sampling methodology;
- A statistical analysis of the hypotheses identified using a standard econometric package; and
- A review of the current methodology for ranking companies with suggestions as to an alternative approach using normalised customer satisfaction scores.

The analysis has been carried out on the basis of data provided by TWUL and publicly available information for the period 2010/11 to 2015/16. The scope of this report does not include any validation of the underlying data.

## 1.3 Structure of this report

The remainder of this report is structured as follows:

- Section 2 sets out the approach and the results of the statistical analysis of CSAT survey data;
- Section 3 reviews the current CSAT methodology and an alternative approach in terms of comparing and ranking companies;
- Appendix A presents the underlying data and methodology for the statistical analysis;
- Appendix B presents the detailed statistical analysis results; and
- Appendix C sets out the details of one potential approach to normalisation methodology.

# 2 Statistical analysis of CSAT survey data

This section sets out the approach, methodology and results of the statistical analysis of CSAT survey scores for the period 2010/11 to 2015/16.

## 2.1 Data sources

A combination of TWUL and publicly available information were used to develop the datasets used to test the various customer, contact and company specific attributes that may influence the CSAT (and hence SIM) scores. The data included:

- Quarterly and weekly CSAT scores for TWUL, at the level of individual customer and associated customer and contact characteristics;
- Quarterly CSAT scores from the "League Tables" across 18 Water-only Companies ("WoCs") and Water and Sewerage Companies ("WaSCs") in England and Wales; and
- Information from the Office of National Statistics ("ONS"), 2011 Census and regional real wages used in Ofwat's 'Basic cost threshold feeder model'.

The TWUL data included:

- Quarterly CSAT scores and associated socio-economic characteristics of respondents and contact attributes for TWUL only. This dataset contains results from six quarterly surveys, one survey per year between 2010 and 2015.
- Weekly CSAT scores and associated contact characteristics for TWUL only, covering the period between October 2014 and November 2015. Each weekly survey samples 200 customers, or more than 10,000 customers during a year.
- League Tables containing the overall quarterly results for all the 18 water companies in England and Wales between 2010 Q1 and 2015 Q4.

The detailed input data used in the econometric analysis is set out in detail in Appendix A.1 and A.2.

## 2.2 Approach and hypotheses tested

The statistical analysis of CSAT survey scores sought to answer the question of whether the CSAT (and hence SIM) scores are affected, in a statistically significant manner, by a number of factors, including customer socio-economic characteristics, company characteristics, customer contact attributes (including the reason for contacting the company), regional differences and the survey duration.

To carry out the statistical analysis, a series of regressions have been run, using a standard econometric package. Due to the nature of the data, a number of different types of econometric models were used, which differed in the detail over which CSAT scores were recorded (customer-level or company-average level), whether the data was available for TWUL only or for other companies, and in terms of the size and the granularity of the dataset itself.

Three types of models, also set out in Table 1 were used to carry out the analysis; it was not possible to use a single type of model to provide an answer to all the questions of interest. For example, it was not possible to use models that used TWUL-only datasets to assess the impact of the type of water company on satisfaction scores. Conversely, it was not possible to use average information for individual companies to assess the impact of reason for

customer contacting company on the satisfaction scores, since this information is not publicly available across water companies in England and Wales.

Quarterly TWUL and Weekly TWUL models use ordered logistic regressions to analyse the survey CSAT scores for Thames only, while Panel Models used panel regressions with random effects to analyse the overall CSAT scores across 18 companies.

**Table 1: Factors investigated through statistical analysis using various models**

| Factor investigated | Quarterly TWUL Models | Weekly TWUL Models | Panel Models |
|---|---|---|---|
| **Customer characteristics** | | | |
| Average weekly income | | | ✓ |
| Average hourly wage | | | ✓ |
| Socio-economic status | ✓ | | ✓ |
| Home ownership | ✓ | ✓ | ✓ |
| Location (rural or urban) | ✓ | | ✓ |
| Age | ✓ | ✓ | ✓ |
| Ethnicity | ✓ | | ✓ |
| Disability | ✓ | | ✓ |
| Qualifications | | ✓ | |
| Proximity to a water treatment works | | ✓ | |
| Gender | ✓ | ✓ | |
| **Contact characteristics** | | | |
| Call or written | | ✓ | |
| Complaint or not | ✓ | ✓ | |
| Resolved or not | ✓ | ✓ | |
| Survey duration | | ✓ | |
| Reason for contact | ✓ | ✓ | |
| **Company characteristics** | | | |
| Type (WoC or WaSC) | | | ✓ |
| Size | | | ✓ |
| Geography | | | ✓ |
| Population turnover | | | ✓ |
| Operating expenditure per customer | | | ✓ |

*Source: Deloitte analysis*

The factors listed in Table 1 were investigated using econometric techniques described in detail in Appendix A. This included a set of model robustness checks to validate the key assumptions underpinning the models.

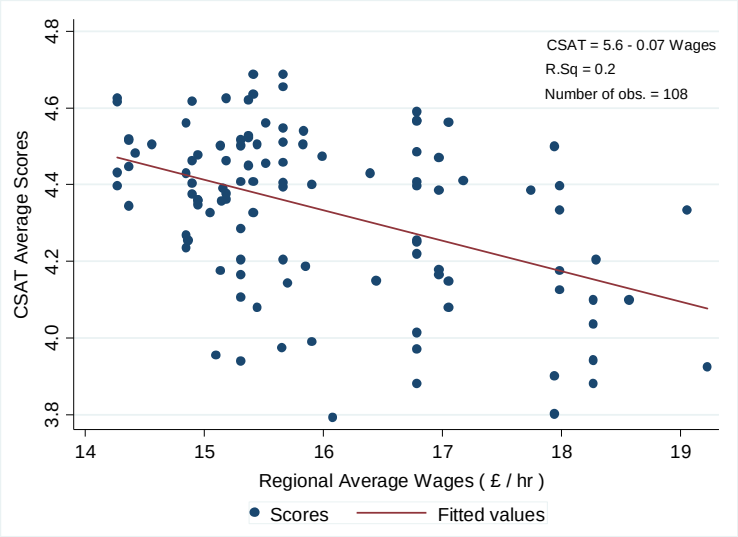
## 2.3 Results of the statistical analysis

The statistical analysis identified a number of factors for which it was possible to reject the hypothesis that these factors had no impact on CSAT scores. These have been grouped below into customer characteristics, contact characteristics and company characteristics. The full set of results can be found in Appendix B.

The key findings are that:

- **Higher customer income or wages in company's region are associated with lower CSAT scores.** This is also the case with other highly correlated variables, including home ownership (as opposed to renting), socio-economic status and household qualifications, although these relationships are not consistent in terms of significance of impact across specifications. The intuition for this result is illustrated in Figure 3, which plots the simple correlation between the regional wage data and CSAT scores. Lower average regional wages are associated with a higher CSAT score. Intuitively, if there was no statistical relationship between CSAT scores and the regional wages, the $R^2$ metric reported in the figure would be expected to be zero. However, the chart also implies that there may be other factors besides wages that are likely to affect the CSAT scores. These factors, such as other company and customer attributes, are examined in detail in the econometric analysis.

**Figure 3: Correlation between CSAT scores and regional wages**




*Source: Deloitte analysis*
*Note: The data reflects annual average CSAT scores from 2010 Q1 to 2015 Q4.*

- **Surveys that take place over a longer period of time are associated with higher CSAT scores.** However, this does not hold when controlling for whether the matter has been resolved, which suggests that longer surveys are associated with higher CSAT scores simply because they are more likely to have been resolved in the meantime.
- **Contacts classified as *not* being a complaint and those that have been resolved at the time of the survey are associated with higher CSAT scores.**
- **Contacts related to water billing are associated with higher CSAT scores**, relative to contacts related to water or waste issues.
- **WoCs are associated with higher CSAT scores compared to WaSCs when controlling for other customer attributes**, but other company characteristics have not been identified as statistically significant (including size, population turnover and operating expenditure by customer).

On the basis of the results of the econometric analysis, the CSAT (and hence SIM) score would be expected to systematically vary across companies independently of water companies' quality of service. As a result, companies may be rewarded and/or penalised under the existing CSAT methodology not just for the quality of service provided to customers, but also for factors that are largely beyond their control, such as the socio-economic characteristics of their population in their appointed area.

In particular, companies whose customers live in areas with lower average income (or wages) would be associated, other things being equal, with a higher CSAT score. Under the current CSAT methodology, Ofwat would be rewarding such companies for the higher CSAT scores. As a result, the water company revenues (and potentially water bills) might increase for those customers who live in areas with lower average income.

The sections below provide further detail on individual customer, contact and company characteristics tested. Across all the tables, the following key is used.

**Table 2: Econometric results: interpretation of the direction of impact results**

| Direction of impact | Interpretation |
|---|---|
| ⇧ | A higher level of the factor is associated with a *higher* CSAT score |
| ⇔ | The factor does not have a statistically significant impact on the CSAT score |
| ⇩ | A higher level of the factor is associated with a *lower* CSAT score |

*Source: Deloitte analysis*

### 2.3.1 Customer characteristics

The results of the econometric estimates of the impact of customer characteristics in relation to the CSAT score is summarised in Table 3. The definitions of the individual characteristics are presented in Appendix A.1.

**Table 3: Econometric results: customer characteristics based on the Q4 2015/16 CSAT survey**

| Customer characteristic | Direction of impact | Interpretation | Statistical significance |
|---|---|---|---|
| Average income (2012/13) | ⇩ | Customers from areas with higher average income are associated with lower SIM scores | Yes (Panel Models) |
| Average wage (2011/12) | ⇩ | Customers from areas with higher average wage are associated with lower SIM scores | Yes (Panel Models) |
| Socio-economic status | ⇩ | Customers with A, B or C status are associated with lower SIM scores compared to customers with D or E status. | Yes (Panel Models) |
| Qualifications | ⇔ | Customers from areas that have a higher share of households with the highest qualifications are associated with lower SIM scores.<br>Conversely, customers from areas that have a higher share of households with no qualifications are associated with higher SIM scores. | No (Weekly Models) |
| Home ownership | ⇩ | Customers from areas with a higher share of home ownership (compared to renting) are associated with lower SIM scores. | Yes (Panel Models) |
| Location (rural or urban) | ⇔ | SIM scores are not affected by whether customers come from predominantly rural or urban areas. | No (Panel Models) |
| Age | ⇔ | Customers' age does not have a statistically significant impact on SIM scores. | Not consistent across models |
| Ethnicity | ⇔ | Customers' ethnicity does not have a statistically significant impact on SIM scores. | Not consistent across models |
| Disability | ⇧ | Customers from areas with lower than average rates of disability per household are associated with lower SIM scores. | Yes (Panel Models) |
| Proximity to a water treatment works | ⇔ | Customers located in specific postcode districts are not associated with SIM scores that are significantly different from other postcodes. | No (Weekly Models) |
| Gender (male) | ⇩ | Responses by male customers are associated with lower SIM scores, compared to responses by female customers. | Yes (Weekly Models) |

*Source: Deloitte analysis*

### 2.3.2 Contact characteristics

The results of the econometric estimates of the impact of contact characteristics in relation to the CSAT score is summarised in Table 4. These results draw primarily on Weekly TWUL Models. The full set of results are set out in Appendix B.

**Table 4: Econometrics results: Contact characteristics**

| Contact characteristic | Direction of impact | Interpretation | Statistical significance |
|---|---|---|---|
| Call contact (written) | ⇩ | Contacts in written form (as opposed to phone contacts) are associated with lower CSAT scores. | Yes |
| Complaint | ⇩ | Contacts categorised as a complaint are associated with lower CSAT scores. | Yes |
| Issue Resolved | ⇧ | Contacts regarding issues that have been resolved at the time of the survey are associated with higher CSAT scores. | Yes |
| Short survey duration | ⇩ | When not controlling for 'Issue Resolved':<br>• Surveys carried out over a shorter period of time are associated with lower CSAT scores. | Yes |
| | ⇔ | When controlling for 'Issue Resolved':<br>• This factor is not statistically significant when controlling for whether the issue has been resolved. | No |
| Reason for contact (water, waste or billing) | ⇩ | Contacts related to **water** are associated with lower CSAT scores, compared to those related to **billing**. | Yes |
| | ⇔ | Contacts related to **waste** are associated with higher CSAT scores compared to those related to billing. However, this result is not consistent across all model specifications. | Not consistent across models |

*Source: Deloitte analysis*

The interpretation of the regression results is in terms of the predicted probabilities for individual customers submitting a given CSAT score. The predicted impacts for selected variables from the Weekly TWUL Models are shown in Table 5. This illustrates the extent to which the probability of a given customer submitting a "Very Satisfied" score differs according to contact attributes such as whether it is categorised as a complaint, whether it has been resolved and what the reason for contacting the water company is.

- For example, if the matter has been resolved at the time of the survey, the respondent is expected to submit a CSAT score of 5, representing the "Very Satisfied" category, with a 70% probability.
- Conversely, if the matter has not been resolved at the time of the survey, the probability of submitting the CSAT score of 5 is materially lower, at 19%.

**Table 5: Illustration of predicted probabilities: contact attributes**

| Variables | Probability of Very satisfied |
|---|---|
| **Reason for contact** | |
| *Billing related* | 57.2% |
| *Waste related* | 58.4% |
| *Water related* | 53.5% |
| **Complaint** | |
| *Yes* | 32.4% |
| *No* | 60.5% |
| **Resolved** | |
| *Yes* | 70.0% |
| *No* | 19.4% |

*Source: Deloitte analysis*

### 2.3.3 Company characteristics

The results of the econometric estimates of the impact of company characteristics in relation to the CSAT scores are summarised in Table 6. These results draw primarily on Panel Models.

**Table 6: Econometric results: Company characteristics**

| Company characteristic | Direction of impact | Interpretation | Statistical significance |
|---|---|---|---|
| Type (WoC) | ⇧ | WoCs tend to be associated with higher SIM scores compared to WaSCs | Yes |
| Size | ⇔ | The size of a water company does not have a statistically significant impact on SIM scores. | No |
| Geography | ⇔ | Company location (S vs NW) is not statistically significant across models that control for income. | No |
| Population turnover | ⇔ | Population turnover in a water company's area does not have a statistically significant impact on SIM scores. | No |
| Company total operating expenditure/customer (2010 to 2015) | ⇔ | Total operating expenditure including third party services and capital maintenance divided by the number of households billed for water services | No |

*Source: Deloitte analysis*

*Note: The company characteristics listed above (with the exception of Company total opex per customer) tend to be correlated, for example the type of company (WoC) and its size. It was not possible to disentangle the effects of the factors in Table 6 fully and the extent to which the type of the water company is associated in a statistically significant manner with the CSAT scores can also capture some of the size and/or geography effect. As a result, the table above needs to be interpreted with caution.*

## 2.4 Limitations and potential further analysis

### 2.4.1 Limitations

**Quarterly and Weekly TWUL Models**

The Quarterly and Weekly TWUL Models reflect company and customer characteristics for a single water company only. As a result, it is not possible to test a number of company characteristics and to assess whether any of these characteristics are associated with statistically different CSAT scores.

Ideally, these models would include qualitatively similar data for a number of companies (although not necessarily all the companies across England and Wales).

**Panel Models**

The panel dataset included in this report contains only the overall weighted average scores for each company and no associated customer or contact characteristics. The panel uses the average CSAT scores by company over time as the dependent variable and a set of explanatory variables indicating differences across regions, companies and their underlying population profiles. There are two limitations associated with this structure:

1. using only the overall company averages reduces the variability of CSAT scores across companies, which in turn limits the precision of the regression coefficients and the tests associated with them; and
2. the majority of the explanatory variables are categorical and highly correlated among each other, thus limiting the potential model specifications tested.

Ideally, a panel would include the quarterly CSAT scores for each company and their associated customer and call attributes. Such a structure would allow the analysis of the impact of company and regional differences on CSAT scores while controlling for changes in scores over time as well as the customer and call attributes. Therefore, there is the possibility that when using a more granular dataset the estimated coefficients for customer and company attributes might be different.

### 2.4.2 Potential further tests

**Sample selection bias**

There are one or two possible sample selection bias issues associated with the current design of the SIM, depending on whether the intent is to capture the satisfaction of all water customers with their supplier, or whether the intent is to only capture the satisfaction of those customers who have made a contact with their water company.

- Currently, CSAT surveys sample customers only from the list of contacts in a given month and not from the wider customer population. Those customers effectively "self-select" to contact the company and if they have different characteristics from the full population of water company customers, then this method of sample selection would create a bias if the intent was to understand the satisfaction levels of all water company customers. For example, customers who are less satisfied with their water company's services may have different socio-demographic characteristics, they may be more likely to contact the company and they may be more likely to subsequently respond to the CSAT survey.
- If, however, the CSAT scores are intended to be representative of only those customers who choose to contact their water company (as is currently the case in the SIM design), then this form of sample selection bias does not arise.

- Nevertheless, sample selection bias can still arise even if the intent of the survey is to only capture the satisfaction levels of customers who have contacted their water company. This could arise for example if customers who respond to the customer survey have different socio-economic characteristics from those who do not respond (and are effectively replaced in the sample by another person). For example, customers who choose to respond to the survey may be systematically more or less dissatisfied customers, or they may disproportionately include those who have spare time to respond to the survey.

Robust and reliable survey results are entirely dependent on the quality of the data sample used. Of particular importance is that each member of the target population of the survey should have equal chances of being selected so that the composition of the sample is representative of the population serviced by each water company.

The econometric analysis of the CSAT scores undertaken in this report shows that different customer attributes are associated with statistically different CSAT scores. Therefore, an empirical test of the sample selection bias may need to be carried out, provided the required data were available.

### Econometric tests

Currently, data on non-respondents is not available.[4] However, these tests could be carried out if in the future, the CSAT surveys tracked customers who either did not complete the survey or chose not to contact the company in the first place.

A binary probit model could be estimated where the dependent variable would indicate which customers responded (a value of 1 would indicate a complete response while a value of 0 would indicate non-response or partial response to the survey) while the set of explanatory variables would include the associated socio-demographic attributes of both respondents and non-respondents.

An additional variable would then be constructed from the predictions of the probit model and included in the CSAT regression.[5] A test of whether this variable is significantly different from zero, is then a test of whether sample selection correction is needed to avoid potential bias.

[4] While theoretically it may be possible to use a Heckman ordered probit procedure to address the issue of non-respondents in some cases, it is not an approach that can be considered as part of this analysis. This is because there is no clear underlying self-selection mechanism that determines why a particular customer does or does not choose to contact their water company, or why they do or do not choose to respond to the survey.
[5] This is known as the inverse Mills ratio.

## 2.5 Sample size

The size of the sample selected is one of the most important features of the overall design of any customer satisfaction survey. The extent to which a measurement of the indicator of interest (in the case of CSAT this is the customers' satisfaction level) is reliable depends on two factors:

- Characteristics of the target customer population covered by each water company; and
- Degree of precision required to identify the satisfaction level.

In general, to achieve a more precise or reliable satisfaction estimate, a larger sample size for the survey is required.

The convention used in many household surveys, including those set by the UN and Eurostat[6] is to choose a sample size that achieves a margin of error between 5% and 10% as well as a 95% confidence level on the indicator to be estimated. The formula used to identify the optimal sample size can be summarised as follows:[7]

$$optimum\ sample\ size = \frac{z^2 * r * (1 - r) * f * k}{p * n * e^2}$$

Where, in the numerator:

- **z** statistic should be 1.96 for the 95% confidence level.
- **r** represents the previous estimates of the target indicator being measured. In this case, this corresponds to the previous proportions of respondents among total respondents categorised in the five customer satisfaction categories in CSAT surveys. For TWUL 2015-16 Q4 survey, the estimated shares of customers are as follows:
  - "Very Satisfied": 55%,
  - "Fairly Satisfied": 19%,
  - "Neither satisfied nor dissatisfied": 9%,
  - "Fairly dissatisfied": 8%; and
  - "Very dissatisfied": 9%.
- **f** represents the sample design effect[8]; this parameter is not included in the sample size formula as the CSAT surveys do not cluster or stratify customers;
- **k** is the multiplier for the anticipated rate of non-response to the survey. This parameter is not relevant for this assessment, since the CSAT surveys are carried out such that the desired sample size is always achieved by design (so there is no need to increase the initial sample to account for expected non-response rate). The parameter has therefore been set to 1.

And in the denominator:

- **p** is the share of the target population in total population. This parameter is equal to 1 as the target population is the set of customers who have contacted their water company (within, effectively, the same set of customers).
- **n** is the average household size. In the numerical example, this is not included since it is not a relevant parameter for the CSAT surveys.
- **e** is the expected margin of error, assumed to be 5%[9] (that is, **e** = 5% * **r**).

[6] Eurostat: Handbook on precision requirements and variance estimation for ESS household surveys.
UN Guidelines: Chapter 3: UN Department of Economic and Social Affairs – Statistics Division. Designing Household Survey Samples: Practical Guidelines, 2005.
[7] This standard formula is usually used to determine the sample size in household surveys.
[8] Design effects account for the difference in sample variability between a simple random sample and a complex sample design such as clustering or stratification in household surveys.
[9] The margin of error refers to the degree of sampling error for the customer satisfaction estimate.

Thus, TWUL-specific information leads to the following optimum sample size formula:

$$optimum\ sample\ size = \frac{z^2 * r * (1 - r)}{e^2}$$

**Table 7: Sample size: illustration for TWUL 2015/16 Q4 survey**

| | Very satisfied | Fairly satisfied | Neither satisfied nor dissatisfied | Fairly dissatisfied | Very dissatisfied |
|---|---|---|---|---|---|
| **Accuracy Error (e)** | 5% | 5% | 5% | 5% | 5% |
| **Previous estimates of the proportion of customers (r)** | 55% | 19% | 9% | 8% | 9% |
| **Optimum sample size** | 380 | 236 | 126 | 113 | 126 |

*Source: Deloitte analysis*

Based on Table 7, the optimum sample size varies depending on which CSAT score (1-5) is of interest. Since in the context of the CSAT surveys, each of the individual categories of customer satisfaction is of interest, the overall sample size for the CSAT surveys needs to be the maximum across the five categories.

In terms of the accuracy error of the 'overall score', selecting the maximum across the five categories implies that the other four categories are estimated with a lower effective error (specific values vary across the four categories). However, an overall score accuracy error is not estimated here.

Therefore, based on the standard methodology used in household surveys, a conventional size of the quarterly CSAT sample size, specifically for TWUL, would include 380 households. This is greater than the current CSAT sample size of around 200 consumers. This result suggests that the sample size used for the Quarterly TWUL Models is smaller than the size estimated above.

However, the choice of the measurement error level in surveys varies across individual studies and in many cases several thresholds are tested.[10] For instance, applying a measurement error of 6.89% to the formula presented above would imply a sample size of 200, consistent with the current size used in CSAT surveys.

The sample size estimated using the approach set out above would vary by company as well as over time, depending on the previous estimates of the proportion of customers falling within the five CSAT categories, although the variation would be limited. Furthermore, based on the formula presented above, the largest sample size (384) would be associated with a 50% proportion of customers in any given category.

[10] See Section 2 in Eurostat guidelines: http://ec.europa.eu/eurostat/documents/3859598/5927001/KS-RA-13-029-EN.PDF

# 3 Review of CSAT methodology

Under the SIM methodology, water companies in England and Wales are currently ranked using 'raw' scores from the CSAT surveys. This relative ranking is then used by Ofwat as the qualitative component of the overall CSAT assessment to determine potential rewards or penalties to individual companies, which in AMP6 can represent up to +6% to -12% of a company's retail household price control revenue.

This section considers an alternative approach to ranking companies, based on a normalised set of customer satisfaction scores. It also sets out the impact of the normalisation of CSAT scores on the relative ranking of water companies using historical CSAT score outcomes.

## 3.1 Drawbacks of direct score comparison

This report considers the use of a normalisation methodology that is used in the customer satisfaction research literature. Academic literature[11] argues that raw scores from customer satisfaction surveys should not be used to directly compare organisations because of the underlying differences in customer expectations across the organisations' respective customer pools.

If customers in different areas indeed have diverging expectations in terms of the quality of service from their water company, then a direct comparison of companies' scores fails to indicate how well companies perform in meeting their own customers' expectations.

Applying this logic to the water sector in England and Wales, if customer expectations differed across individual WoCs and WaSCs, may mean that a normalised score comparison could be more appropriate than a direct score comparison. This would be the case particularly if, as shown in Section 2, there are indeed differences in customers' socio-economic characteristics across individual companies and these characteristics have a statistically significant impact on the CSAT scores.

## 3.2 Normalisation of customer satisfaction scores

There are a number of potential approaches that can be taken to normalise customer satisfaction scores, with the preferred approach being dependent on the data availability and the characteristics of the target population being observed.

### 3.2.1 An approach based on econometric analysis

The preferred approach would involve using a variant of the Weekly TWUL econometric model described in Appendix B. Such a model, not included in this report, would effectively extend the TWUL-only model to include all other WoCs and WaSCs to control for both individual and company characteristics. While currently CSAT survey results at the customer level are not available for the rest of the companies, such a model would be able to better approximate the scores across companies while controlling for differences in service quality across companies and variations in customer expectations.

[11] Recent economic literature proposes normalisation techniques of customer satisfaction scores from public transport surveys that aim to compare public transport operators across regions, such as: Trompet et al. (2013), Johnson et al. (2002), TCRP Report 141 (2010).

For example, a variant of the model would include individual customer characteristics (income, age, family size) and variables that differ by company but are common to all the customers (specific services). Such a model specification would enable a comparison of how customers in different areas would rate a specific water company if those customers had different socio-economic characteristics.

### 3.2.2 An approach without econometrics

A normalisation approach to CSAT scores helps address the inherent differences in customer expectations in different regions by standardising the components of the satisfaction score (related to billing, water and waste).

Since customers' expectations are not observable, an approach without econometrics indirectly uses customers' satisfaction with other services from the same company as a benchmark when evaluating customers' satisfaction with the service of interest. For example, the "raw" satisfaction scores on water and waste can be used as a benchmark when assessing whether the "raw" satisfaction score on billing is high or low for a particular company. The details of this methodology are set out in Appendix C.

The limitation of this normalisation methodology is that it implicitly removes any company-specific factors that impact all services equally. For example, if a specific water company received low rating across all services provided, the normalisation methodology would discard this information. As a result, the methodology fails to distinguish whether the relatively lower scores for a company are driven by the underlying customers' expectations, or whether they are driven by a relatively lower quality of the services provided.

### 3.2.3 A simplified approach (an illustration)

The approach set out in Section 3.2.1 is the preferred approach to normalising CSAT scores, since it makes the best use of the data and the econometric analysis carried out in relation to customer satisfaction.

However, in the absence of the required data inputs as set out in Section 3.2.1, it is possible to develop the intuition for this approach by considering a simplified approach that illustrates its logic. This analysis is not intended to serve as an estimate of the actual magnitude of the potential changes in the CSAT scores (or in company rankings). Rather, the objective is to develop the intuition behind the normalisation technique and to encourage the reader to consider whether further analysis set out in Section 3.2.1 could be warranted.

This illustration focuses on Panel Model results presented in Appendix B and particularly on one of the two explanatory variables that varies over time and that is statistically significant, namely the average income.[12] As shown in the Appendix, average income across regions have a negative and statistically significant impact on CSAT scores. These specifications provide an illustration of the impact that changes in income are estimated to have on CSAT scores.

The average coefficient estimate for the income variable across the panel models 15 to 24 presented in Table 25 is -0.056. This estimate was used to 'normalise' the raw CSAT scores by considering the distance between each company's explanatory variable (average income in the company's area) and the company average. Specifically:

- First, an 'adjustment factor' for each CSAT score was computed as the difference between each company's income observation for a given quarter and the corresponding quarterly average income multiplied by the income coefficient estimate; and

[12] It would be appropriate for the econometric model to control for other company or customer related variables that can influence the scores, but these are not available on a time-variant basis. Additional variables may include for example variations in the quality of services offered or customer expectations across companies, but these were not available for this study.

- Second, the adjustment factor was added to each CSAT score across quarters and companies.

Table 8 compares the rankings of companies using the raw and income adjusted CSAT scores from the 2015 Q4 survey. For companies in regions with relatively low income (compared to the company average), this technique tends to reduce the normalised CSAT score, while for companies in regions with relatively high income it tends to increase the normalised CSAT score. The results of the normalisation indicate that companies located in regions with relatively higher income, tend to improve their ranking, depending on the change in income.

**Table 8: Ranking of raw and income adjusted CSAT scores for 2015 Q4**

| Rank | 2015 Q4 | |
|---|---|---|
| | Raw scores | Scores adjusted for income effect |
| 1 | Bournemouth (4.61) | Portsmouth (4.65) |
| 2 | Portsmouth (4.6) | Bournemouth (4.6) |
| 3 | Dee Valley (4.59) | Dee Valley (4.51) |
| 4 | Welsh (4.52) | Wessex (4.47) |
| 5 | Wessex (4.5) | Yorkshire (4.45) |
| 6 | Yorkshire (4.49) | Bristol (4.44) |
| 7 | Bristol (4.47) | Welsh (4.43) |
| 8 | South Staffs (4.45) | South Staffs (4.43) |
| 9 | Severn Trent (4.42) | South East (4.38) |

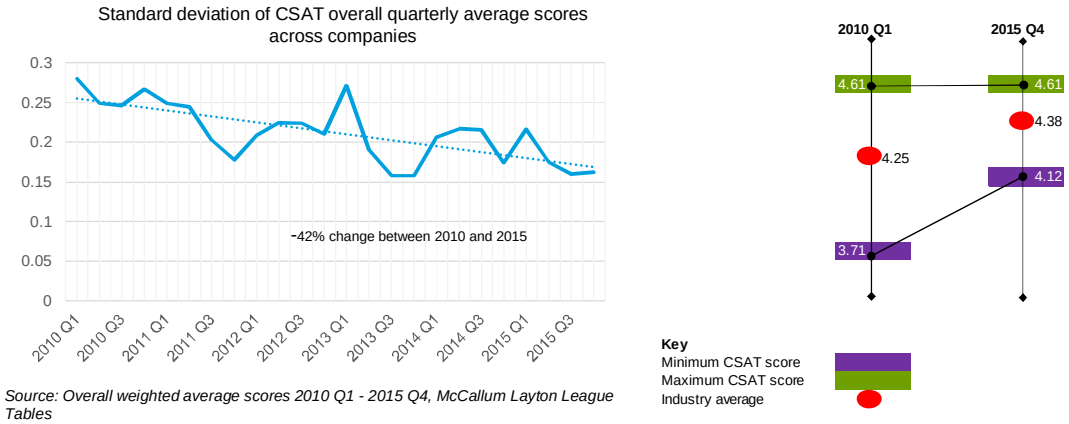
| Rank | 2015 Q4 | |
|---|---|---|
| | Raw scores | Scores adjusted for income effect |
| 10 | Anglian (4.38) | Affinity (4.37) |
| 11 | United Utilities (4.36) | Severn Trent (4.37) |
| 12 | South East (4.33) | Anglian (4.37) |
| 13 | Affinity (4.31) | United Utilities (4.31) |
| 14 | Northumbrian (4.28) | Sutton (4.27) |
| 15 | South West (4.18) | Thames (4.26) |
| 16 | Sutton (4.15) | Northumbrian (4.23) |
| 17 | Southern (4.13) | Southern (4.18) |
| 18 | Thames (4.12) | South West (4.15) |

*Source: Deloitte analysis*

## 3.3 Convergence of CSAT scores over time

The variability in CSAT survey scores across companies has been declining over time. Figure 4 below shows the change in the scores' standard deviation across overall company averages. Between 2010 Q1 and 2015 Q4 the standard deviation of the CSAT surveys declined by 42%.

**Figure 4: CSAT standard deviation over time**




*Source: Overall weighted average scores 2010 Q1 - 2015 Q4, McCallum Layton League Tables*

*Source: Deloitte analysis*

The decrease in the standard deviation of scores around the industry average implies a disproportionate impact of penalties (either in AMP5 or AMP6) for the water companies ranked below the industry average, as the distance between the lowest and highest ranked companies has declined over time. For instance, in 2010 Q1, a one-standard deviation below the industry average implied a difference in scores of 0.28 units (in terms of the CSAT score) while in 2015 Q4 the difference dropped to 0.16 units.

However, the range of penalties and rewards applied to companies for being ranked below and above the industry average in 2015 was the same as in 2010. As a result, the same penalties and rewards are now applied to a narrower distribution of customer satisfaction scores, which means that the penalties or rewards are now more sharply pronounced relatively to the customer satisfaction outcomes.

# Appendix A
# Methodology – econometric analysis

Appendix A presents the data sources and the econometric techniques that were used to carry out the statistical analysis of CSAT survey data.

## A.1 Data sources

Deloitte used a combination of TWUL and publicly available information to develop the datasets used to test the various customer, contact and company attributes that may influence the CSAT scores.

The TWUL data included:

- Quarterly CSAT scores and associated socio-economic characteristics of respondents and contact attributes for TWUL only. This dataset contains results from six quarterly surveys, one survey per year between 2010 and 2015.
- Weekly CSAT scores and associated contact characteristics for TWUL only, covering the period between October 2014 and November 2015. Each weekly survey samples 200 customers, or more than 10,000 customers during a year.
- League Tables containing the overall quarterly results for all the 18 water companies in England and Wales between 2010 Q1 and 2015 Q4.
- Company operating expenditure by customer, based on published regulatory accounts.

In addition, Deloitte received postcode information from TWUL regarding the location of a specific water treatment plant, Mogden Sewage Treatment Works ("Mogden"). This information was used in the weekly data analysis to construct an additional variable that took on a value of 1 for customers located in the vicinity of Mogden and 0 for all other customers.

Deloitte has collected publicly available information on average socio-economic population characteristics across water companies, including from the Office for National Statistics ("ONS")[13], Ofwat[14] and 2011 Census[15] data. This included:

- **Average age.** The 2011 Census data were provided at the postcode district level and represented the mean age of all residents.
- **Tenure.** The 2011 Census data was provided at the level of postcode district and included the share of households who owned or privately rented their accommodation.
- **Qualifications.** The 2011 Census data was provided at the postcode district level and included the share of the population, among residents aged 16 and over, with no qualifications and the share with the highest qualification, defined as Level 4 qualifications and above.
- **Income.** The data from ONS included Net weekly household income for England and Wales for 2011/12, by middle layer super output area ("MSOA"). The data from Ofwat included real regional wage water (in £/hour) for 2012/13, by individual water companies. The most recent data, from 2009/10 to 2012/13 was used in this report.

[13] ONS 'Small area income estimates': http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-416744.
[14] Ofwat real regional wages used in the "Basic cost threshold feeder model": http://webarchive.nationalarchives.gov.uk/20150624091829/https://www.ofwat.gov.uk/pricereview/pr14/wholesale/prs_web140404pr14wholesalecostasses
[15] Census tables are available at: http://www.nomisweb.co.uk/census/2011/data_finder.

Table 9 presents a summary for each of the datasets used in the analysis.

**Table 9: Data sources used in the econometric analysis**

| Data description | Variables included | Number of observations |
|---|---|---|
| Six quarterly CSAT surveys tracking TWUL customers only, one survey per year between 2010 and 2015 | • Contact characteristics<br>• Customer socio-economic attributes | Around 200 respondents per quarterly survey, for a total of about 1000 observations. |
| Weekly CSAT scores for TWUL customers only covering the period October 2014 - November 2015 | • Contact (call) characteristics<br>• Survey duration<br>• Respondents' gender<br>• Socio-economic attributes at the postcode district level from the 2011 Census<br>• Proximity to Mogden | Around 200 respondents per weekly survey |
| League Tables containing CSAT scores for each of the 18 water companies covering the period 2010 Q1 through to 2015 Q4 | • Socio-economic attributes at the regional level from TWUL (population turnover, socio-economic status, age, disability, tenure, location and ethnicity)<br>• Net household weekly income estimate for 2011/12 from the ONS<br>• Regional real wages (from Ofwat's "Basic cost threshold feeder model") for 2009/10 to 2012/13.<br>• Company attributes from TWUL data, including size, region, type (WoC/WaSC) | 18 observations per quarterly survey for a total of 432 observations |

*Source: Deloitte analysis*

## A.2 Data manipulation

### A.2.1 Quarterly TWUL Models

Quarterly CSAT scores for TWUL between 2010 and 2015 contain customer-level CSAT scores and their associated characteristics as well as contact attributes. While each quarterly survey contains responses from 200 customers, entries for those who responded "Don't know" or "Refused" to any of the questions asked have been removed. In addition, all business respondents have been removed and only domestic customers have been retained.

Responses to all the questions, except the satisfaction scores, were recorded as categorical variables. Because of the low number of observations for some of the categories, the following variables have been re-coded:

- *Tenure* has been merged to include only two categories: owned (includes part-owned) and not owned (includes privately rented accommodation as well as accommodation rented from local authority or housing association);
- *Status: categories A and B have been aggregated due to the low number of observations for A;*
- *Age* aggregates the responses for "Under 25" and "25 – 34" into "Under 35" due to the low number of observations recorded for the "Under 25" category; and
- *Ethnicity* is categorized between "White British" and "Other".

Table 10 and Table 11 present the customer and call variables included in the analysis.

**Table 10: Customer attributes (Quarterly TWUL Models)**

| Variable | Categories | Variable labels in the models |
|---|---|---|
| **Tenure** | Owned<br>Not owned | tenure |
| **Age** | Under 35<br>35 - 44<br>45 - 54<br>55 -64<br>65+ | age |
| **Location** | Large city<br>Small town<br>Village | location |
| **Disability in the household** | Yes<br>No | disability |
| **Status** | AB<br>C1<br>C2<br>D<br>E | status |
| **Ethnicity** | White<br>Other | ethnicity |
| **Gender** | Male<br>Female | gender |

*Source: TWUL*

**Table 11: Call attributes (Quarterly TWUL Models)**

| Variable | Categories | Variable labels in the models |
|---|---|---|
| **Is the contact categorised as a complaint?** | Yes<br>No | complaint |
| **Has the matter been resolved?** | Yes<br>No | resolved |
| **Reason for contact** | Billing<br>Water<br>Waste | reason |

*Source: TWUL*

### A.2.2 Weekly TWUL Models

Weekly CSAT scores come from TWUL Weekly Tracker surveys, which contain only the contact attributes and the postcode district of the respondents, but not all the respondents' socio-economic characteristics that are included in the Quarterly TWUL models. The data was augmented using the 2011 Census data at the postcode district level as shown in Table 13. The weekly data used in the econometric analysis are shown in Table 12 and Table 13.

**Table 12: Customer attributes (Weekly TWUL Models)**

| Variable | Categories | Variable labels in the models |
|---|---|---|
| **Average age within the postcode district** | Continuous variable | average_age |
| **Share of residents with highest qualification (L4)** | Continuous variable | highest_qualification |
| **Share of residents with no qualification within the postcode district** | Continuous variable | no_qualifications |
| **Share of owned residences within the postcode district** | Continuous variable | tenure_share_owned |
| **Share of rented residences within the postcode district** | Continuous variable | tenure_share_rented |
| **Gender** | 1 = Male<br>0 = Female | gender |
| **Proximity to Mogden** | 1 = yes<br>0 = no | odour |

*Source: TWUL, ONS, 2011 Census, Department for Communities and Local Government*

**Table 13: Contact attributes (Weekly TWUL Models)**

| Variable | Categories | Variable labels in the models |
|---|---|---|
| **Is the contact categorised as a complaint?** | Yes<br>No | contact_complaint |
| **Reason for contact (aggregated)** | Billing<br>Water<br>Waste | main_reason |
| **Method of contact** | Written<br>Telephone | contact |
| **Survey duration** | 1 = 1-week survey<br>2 = 2-week survey | long_survey |
| **Has the matter been resolved?** | Yes<br>No | resolved |

*Source: TWUL, ONS, 2011 Census*

### A.2.3 Panel Models

The overall CSAT scores at the company level have been obtained from TWUL in the format of McCallum Layton League Tables and contain the weighted overall CSAT scores for each company over the period between 2010 Q1 and 2015 Q4.

For each quarter the data includes 18 scores (one for each WoC and WaSC), or 432 total observations. These overall scores per company have been combined over time into a panel, and have been augmented with data from the ONS, Ofwat and the 2011 Census. Table 15 presents the regional and company attributes included in the analysis.

**Table 14: Company, customer attributes (Panel Models)**

| Variable | Categories | Variable labels in the models |
|---|---|---|
| **Real wage (2012/13) by company** | HW (high wage)[16]<br>LW (low wage) | ofwat_income |
| **Real wages (2009/10 – 2012/13) by company** | Continuous variable[17] | inc |
| **Net weekly income by company region (2012 values)** | H (higher than sample average)[18]<br>L (lower than sample average) | ons_income |
| **Predominant socioeconomic status of the population** | AC (status: A, B, C1, C2)<br>DE (status D and E) | status |
| **Age structure of the population** | younger[19]<br>older | age |
| **Share of homeowners for each company area** | share_high[20]<br>share_low | tenure |
| **Location** | urban<br>rural[21] | location |
| **Ethnicity** | higher[22]<br>lower | ethnicity |
| **Disability** | Higher[23]<br>Lower | disability |
| **Population turnover rate** | H (higher than sample average)<br>L (lower than sample average) | turnover |
| **Geographical location** | North West (NW)[24]<br>South (S) | company_geography |
| **Company operating expenditure by customer** | Continuous variable[25] | exp4 |
| **Size of the company** | large (L)[26]<br>small (S) | company_size |
| **Type of water company** | WoC<br>WaSC | company_type |

*Source: TWUL, ONS, 2011 Census*

[16] Companies have been categorised as "HW" if the average wage across the company's customers is higher than the average wage across all companies. This data included Welsh Water.
[17] Regional annual real wages (from Ofwat's "Basic cost threshold feeder model") were disaggregated at quarterly intervals.
[18] Companies have been categorised as "L" if the average net weekly income was below the overall average across companies. This excludes Dee Valley.
[19] Companies have been categorised as "younger" if the share of residents under 54 years old for a given company is higher than the share of people aged above 55.
[20] Companies have been categorised as "share_low" if the share is lower than the average homeownership across all companies.
[21] Companies have been categorised as "rural" if the share of residents living in a village is higher than the share of people living in a city.
[22] Companies have been categorised as "higher" if the share of residents who categorised themselves as "White: British" for a particular company area is higher than the average across all company regions.
[23] Companies have been categorised as "higher" if the average share of households with disabilities is higher than the overall average rate of households with disabilities across companies.
[24] "NW" category includes Welsh Water, Yorkshire, South Staffs, Northumbrian, United Utilities and Severn Trent. The remaining WoCs and WaSCs are categorised as "S".
[25] Total company operating expenditure per customer was retrieved from the annual regulatory accounts between 2010 and 2015. The total operating expenditure, as defined in this report, include: total operating expenditure, capital maintenance and third party services for wholesale water-only services. The number of customers is measured by the number of households billed for water services.
[26] Companies have been labelled as "L" for their size if the total population of their respective areas was lower than the average population across all the company areas. Higher populations than average were categorised as "H".

The ONS net weekly income data were aggregated from the local authority (LAD) level to the company level by using population weights for each LAD. In addition, where LADs are serviced by both a WoC and a WaSC, these areas were duplicated in the dataset such that their related scores and income estimates are counted in the company average. For example, in Barnet, Thames provides sewerage services while Affinity provides water services.

## A.3 Econometric models

Deloitte estimated two types of econometric models to analyse the data described in Sections A.1 and A.2. Ordered logit models were used for the Quarterly and Weekly models covering TWUL customers only and a Panel model was used to analyse the variation in scores across companies over time.

### A.3.1 Ordered logistic models (Quarterly and Weekly models)

The ordered logistic regressions seek to identify an association between the CSAT scores and the socio-economic and contact attributes. The CSAT scores have five possible values ordered on a scale from 1 to 5, where 1 = Very dissatisfied and 5 = Very satisfied. The ordered categories of the scores can be analysed with two techniques: ordered probit and ordered logistic regressions.

- Probit regressions assume the disturbance terms to be normally distributed, however, the CSAT scores tend to be highly skewed towards the higher end of the scale, which raises concerns with the tail part of the probability distribution curve of the disturbance terms.
- Logistic regressions assume that the random disturbance terms of the scores follow a standard logistic distribution, which has a similar shape as the normal distribution, but heavier tails, allowing for skewed observations. As a result, logistic regressions are more appropriate for this analysis.

Ordered logistic distributions are also known as proportional odds models. This means that the estimated coefficients that describe the relationship between the five levels of the CSAT scores and the various explanatory variables included in the regressions are the same at all CSAT satisfaction levels. In other words, the relationship between the explanatory variables and the lowest score (very dissatisfied) is exactly the same as the relationship between the highest score (very satisfied) and the explanatory variables. The proportional odds assumption has been tested, using a Brant test of parallel regression assumption ("Brant test"), for all the ordered logistic regressions reported in this report.

Appendix B presents in detail the results of ordered logistic regressions for the weekly and quarterly datasets, together with diagnostic and robustness checks and output interpretation. The same procedures have been used to analyse the quarterly CSAT scores, although only the estimated outputs have been reported.

### A.3.2 Panel models

The League Tables provided by McCallum Layton contain the overall weighted average CSAT scores for each of the 18 water companies in England and Wales for each quarter covering the period 2010 Q1 through to 2015 Q4. The structure of the overall scores, where the scores for each company are observed over time, can be analysed as a panel structure. Panel data control for effects that are either unobserved or hard to measure, such as regional differences or differences in the quality of service across companies.

The conventional approach to modelling panel data, is to allow the constant term in the model to vary randomly across individuals (in this case water companies). The fundamental distinction, is then whether to treat these individual effects as fixed or random.

- In the fixed effects model, the constant (company effect) is correlated with the explanatory variables, thus allowing for a limited form of endogeneity. Although it is possible to

accommodate this endogeneity, it is not possible to estimate the effects of time-invariant variables
- In the random effects model, it is assumed that the constant is purely random. The advantages of this model are that it permits estimator of time varying and time invariant regressors.

**Hausman test**

All the panel model specifications included in the analysis are estimated with random-effects. The random-effects specification assumes that the heterogeneity across water companies is random and uncorrelated with the included explanatory variables. The choice of this specification has been motivated by the large number of time-invariant explanatory variables (only two explanatory variables are time-varying) and a series of statistical tests that compare the random-effects specification with fixed-effects and ordinary least-squares (OLS).

The choice between random and fixed-effects was investigated by running a Hausman test. This latter tests the difference between the fixed and random effects estimators and if there is no significant difference between the two, then the more efficient random-effects specification is chosen. The Hausman test results for model specifications that include the time-varying variables show that the random-effects specification should be chosen.

For example, the Hausman results for Panel Model 23 presented in Table 25 is presented below and it shows that the null hypothesis of no differences between the fixed and random specification is not rejected. In other words, the random-effects specification should be chosen. Panel model 23 was chosen as an illustration as it contains the two time-varying explanatory variables available, real wages by company region and company expenditures.

**Table 15: Hausman test for Panel Model 23**

| | | **Estimated Coefficients** | | |
|---|---|---|---|---|
| Explanatory variables | Fixed effects | Random effects | Difference | Std. Error |
| Time trend | 0.0021 | 0.0013 | 0.0008 | 0.0011 |
| Real wages | -0.0327 | -0.0482 | 0.0155 | 0.0222 |
| Company expenditures | 0.0012 | 0.0012 | 0.0000 | 0.0008 |
| Test: Ho: difference in coefficients not systematic | | | | |
| Chi2(3) = 1.48 | | | | |
| **Prob>chi2 = 0.6880** → Random-effects specification should be chosen. | | | | |

*Source: Deloitte analysis*

It should be noted that for the panel model specifications that include only time-invariant variables, such as the models presented in Table 23 and Table 24, the Hausman test cannot be conducted. Therefore, the random-effects models should also be tested against a simple OLS specification. This was tested with a Breusch-Pagan LM test. The test was conducted for all the panel models and the results for all of them show that the random-effects specification is appropriate.

In addition to the statistical tests discussed above, correlations between the continuous variables included in the panel models and the random-errors component of the residuals for each model are presented in Table 16, based on bivariate regressions between the residuals and the selected variables. The estimated correlation coefficients for all the models are not statistically significant, thus validating the random-effects assumption, which requires the covariates to be uncorrelated with the company random-effects. While this is not a full econometrics test, it provides an illustration of the underlying relationships between the variables.

## Table 16: Correlations between the random-error component of the residuals and continuous covariates

| Correlation coefficients | Panel Model 15 | Panel Model 16 | Panel Model 17 | Panel Model 18 | Panel Model 19 |
|---|---|---|---|---|---|
| Real wages | -0.231 | -0.222 | -0.192 | -0.231 | -0.234 |
| | (0.2) | (0.205) | (0.198) | (0.217) | (0.204) |
| *95% confidence interval* | *[-0.567 0.199]* | *[-0.568 0.216]* | *[-0.522 0.238]* | *[-0.595 0.233]* | *[-0.573 0.209]* |
| **Observations** | **428** | **428** | **428** | **428** | **428** |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

| Correlation coefficients | Panel Model 20 | Panel Model 21 | Panel Model 22 | Panel Model 23 | Panel Model 24 |
|---|---|---|---|---|---|
| Real wages | -0.221 | -0.235 | -0.204 | -0.142 | -0.222 |
| | (0.203) | (0.204) | (0.198) | (0.223) | (0.204) |
| *95% confidence interval* | *[-0.562 0.206]* | *[-0.575 0.208]* | *[-0.529 0.229]* | *[-0.570 0.297]* | *[-0.563 0.219]* |
| Company expenditures | | | | -0.14 | |
| | | | | (0.188) | |
| *95% confidence interval* | | | | *[-0.470 0.239]* | |
| **Observations** | **428** | **428** | **428** | **428** | **428** |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

*Source: Deloitte analysis*

*Note: The sampling distributions used to compute the standard errors associated with each correlation coefficient were approximated by re-sampling the original data for each model (a technique known as bootstrapping).*

# Appendix B
# Detailed modelling results

This Appendix sets out the detailed modelling results from three types of models presented in Appendix A, covering the Quarterly TWUL Models, Weekly TWUL Models and Panel Models.

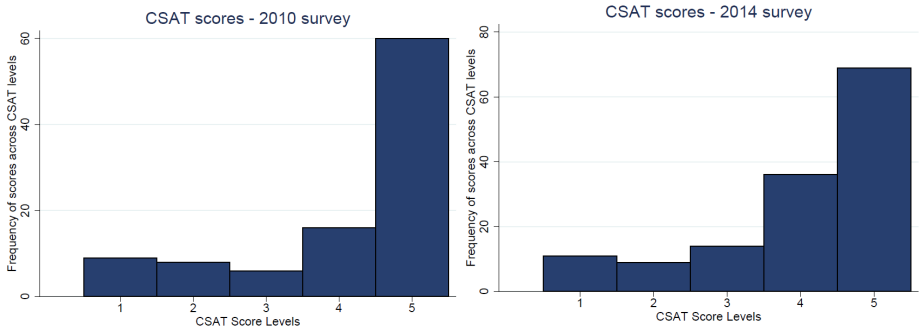
## B.1 Quarterly TWUL Models

### B.1.1 Data characteristics

There are two key features of the data underpinning the Quarterly TWUL Models that impact on the regression results:

1. The low number of observations for each quarterly survey (typically around 200 per survey)
2. The low variability in CSAT scores (see Figure 5).

**Figure 5: Variability in CSAT scores (Quarterly TWUL Models), illustration for 2010 and 2014**




*Source: McCallum Layton Tables provided by TWUL*

In the data used in this report, the low variability of the dependent variable often results in statistically insignificant coefficient estimates, particularly for respondents' socio-economic characteristics. Between 60% and 75% of the scores are in categories 4 and 5 across the CSAT surveys between 2010/11 and 2015/16. Contact characteristics, such as complaints and the timing of resolutions tend to be statistically significant across the majority of model specifications estimated.

In addition, given the low number of observations and the high correlation among the explanatory (categorical) variables, it is not possible to test the proportional odds assumption (see Section A.3.1) for models that include more than two or three categorical variables for each model specification.

### B.1.2 Econometric results

Across a number of models, the following key results have been identified:

- **Resolution.** Respondents whose issue had been resolved at the time of the survey are associated with a higher CSAT score.
- **Complaint.** Respondents whose engagement with TWUL was categorised as a complaint are associated with a lower CSAT score.
- **Reason for contact.** Respondents whose engagement with TWUL relates to water or waste (as opposed to billing) are associated with a lower CSAT score, although this result is not statistically significant.
  **Socio-economic characteristics.** Respondents' attributes were typically not identified as having a statistically significant impact on the reported CSAT scores. In cases where these were found as being statistically significant (for example, some categories of age indicated that higher age bracket was associated with a higher score), this was not consistently the case over time or the sign of the impact changed.

Table 17 summarizes the results from model specifications that include the contact characteristics and customer attributes. The output includes the coefficient estimates, their z-tests and their associated p-values. The variables describing the contact characteristics, Complaint and Resolved, are statistically significant while the majority of the socio-economic variables are not (or this is not consistent across model specifications).

As an example of interpreting the results, the negative coefficient for "Complaint" (Yes) means that the lower CSAT scores are associated with contacts that are categorized as a complaint as opposed to not being categorised as such. Similarly, the positive coefficient of "Resolved" (Yes) implies that matters that have been categorised as resolved at the time of the survey are associated with a higher CSAT score.

However, for these models the Brant test (described in section A.3.1) could not be conducted due to the lack of a sufficient number of observations. In addition, as described in Section 2.5, the sample size of 200 customers per survey is also lower than what conventional sample size estimate would be (380 customers for TWUL). These results therefore are intended to provide only an indication of the potential customer attributes' impact on CSAT scores.

Further below, Table 18 presents a set of results for a selected set of explanatory variables (dropping those that have not been found to be significant in Table 17) for six models across the 2010/11 through to 2015/16 surveys.

## Table 17: Ordered logistic models (Quarterly TWUL Models) Set 1

**Dependent variable: SIM scores from 2014-15 survey**

| Explanatory Variables | Quarterly 2014-2015 (1) | Quarterly 2014-2015 (2) | Quarterly 2014-2015 (3) | Quarterly 2014-2015 (4) | Quarterly 2014-2015 (5) | Quarterly 2014-2015 (6) | Quarterly 2014-2015 (7) | Quarterly 2014-2015 (8) | Quarterly 2014-2015 (9) | Quarterly 2014-2015 (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| **Complaint** | | | | | | | | | | |
| *Yes* | -1.648*** | -1.781*** | | -1.715*** | -1.798*** | -1.671*** | -1.767*** | -1.852*** | -2.002*** | |
| | (0.449) | (0.465) | | (0.443) | (0.465) | (0.445) | (0.475) | (0.452) | (0.47) | |
| **Resolved** | | | | | | | | | | |
| *Yes* | 2.048*** | 2.043*** | 2.171*** | 1.957*** | 2.050*** | 2.028*** | 2.028*** | | | 2.100*** |
| | (0.396) | (0.394) | (0.398) | (0.388) | (0.394) | (0.391) | (0.4) | | | (0.39) |
| **Tenure** | | | | | | | | | | |
| *Owned* | -0.4 | | | -0.0771 | | | | | | |
| | (0.442) | | | (0.442) | | | | | | |
| **Gender** | | | | | | | | | | |
| *Male* | 0.305 | 0.193 | 0.27 | | 0.17 | 0.306 | 0.0949 | | | |
| | (0.363) | (0.357) | (0.372) | | (0.358) | (0.347) | (0.371) | | | |
| **Age** | | | | | | | | | | |
| *Under 35* | 0.105 | | 0.287 | | | | | 0.208 | 0.154 | |
| | (0.501) | | (0.486) | | | | | (0.481) | (0.495) | |
| *45-54* | 1.003* | | 0.874 | | | | | 0.85 | 0.758 | |
| | (0.575) | | (0.567) | | | | | (0.548) | (0.554) | |
| *55-64* | 0.521 | | 0.581 | | | | | 0.00459 | -0.097 | |
| | (0.566) | | (0.582) | | | | | (0.527) | (0.541) | |
| *65+* | 0.725 | | 0.727 | | | | | 0.52 | 0.66 | |
| | (0.507) | | (0.5) | | | | | (0.511) | (0.524) | |
| **Reason for contact** | | | | | | | | | | |
| *Water related* | | -0.103 | -0.446 | | -0.112 | | -0.128 | -0.581 | -0.561 | -0.323 |
| | | (0.432) | (0.459) | | (0.434) | | (0.446) | (0.432) | (0.455) | (0.443) |
| *Waste related* | | 0.325 | -0.0334 | | 0.33 | | 0.309 | 0.00349 | -0.00774 | 0.00843 |
| | | (0.47) | (0.489) | | (0.472) | | (0.486) | (0.451) | (0.475) | (0.476) |
| **Location** | | | | | | | | | | |
| *Small/medium town* | | 0.354 | | | 0.362 | | 0.212 | | | 0.369 |
| | | (0.431) | | | (0.434) | | (0.444) | | | (0.44) |
| *Village* | | -0.394 | | | -0.428 | | -0.35 | | | -0.183 |
| | | (0.498) | | | (0.501) | | (0.512) | | | (0.501) |
| **Status** | | | | | | | | | | |
| *C1* | | | -0.00738 | | | | 0.182 | | 0.575 | -0.124 |
| | | | (0.399) | | | | (0.414) | | (0.405) | (0.402) |
| *C2* | | | 1.240* | | | | 1.348* | | 1.513** | 1.290* |
| | | | (0.708) | | | | (0.75) | | (0.693) | (0.701) |
| *D* | | | 0.953 | | | | 0.954 | | 1.204 | 0.818 |
| | | | (0.836) | | | | (0.842) | | (0.821) | (0.82) |
| *E* | | | 0.214 | | | | 0.00702 | | 0.0938 | -0.016 |
| | | | (0.688) | | | | (0.678) | | (0.65) | (0.704) |
| **Ethnicity** | | | | | | | | | | |
| *White: British* | | | | 0.00467 | | -0.029 | | | -0.137 | |
| | | | | (0.415) | | (0.383) | | | (0.414) | |
| **Disability** | | | | 0.413 | 0.496 | 0.396 | | 0.172 | | 0.362 |
| *Yes* | | | | (0.61) | (0.618) | (0.61) | | (0.604) | | (0.589) |
| Constant cut1 | -1.693*** | -1.758*** | -0.838 | -2.021*** | -1.743*** | -1.773*** | -1.646*** | -2.969*** | -2.714*** | -1.343** |
| | (0.637) | (0.563) | (0.653) | (0.571) | (0.566) | (0.569) | (0.623) | (0.524) | (0.618) | (0.565) |
| Constant cut2 | -0.801 | -0.891* | -0.0576 | -1.135** | -0.87 | -0.882* | -0.767 | -2.217*** | -1.933*** | -0.574 |
| | (0.598) | (0.53) | (0.634) | (0.526) | (0.532) | (0.529) | (0.598) | (0.475) | (0.578) | (0.541) |
| Constant cut3 | 0.0614 | -0.0625 | 0.739 | -0.292 | -0.036 | -0.0396 | 0.075 | -1.477*** | -1.167** | 0.199 |
| | (0.587) | (0.519) | (0.637) | (0.506) | (0.521) | (0.51) | (0.591) | (0.451) | (0.56) | (0.538) |
| Constant cut4 | 1.579*** | 1.450*** | 2.213*** | 1.180** | 1.479*** | 1.438*** | 1.620*** | -0.178 | 0.184 | 1.642*** |
| | (0.607) | (0.536) | (0.663) | (0.517) | (0.538) | (0.525) | (0.606) | (0.436) | (0.551) | (0.558) |
| **Observations** | **139** | **139** | **139** | **139** | **139** | **139** | **139** | **139** | **139** | **139** |
| **Pseudo R2** | 0.1440 | 0.1410 | 0.1233 | 0.1310 | 0.1428 | 0.1331 | 0.1544 | 0.0659 | 0.0858 | 0.1151 |
| **LR chi2(9)** | 52.3 | 51.22 | 44.78 | 47.59 | 51.89 | 48.34 | 56.08 | 23.93 | 31.18 | 41.81 |
| **Prob > chi2** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

*Source: Deloitte analysis*

## Table 18: Ordered logistic models (Quarterly TWUL Models) Set 2

**Dependent variable: SIM scores from quarterly surveys between 2010-11 and 2015-16**

| Explanatory Variables | Quarterly 2010 Model | Quarterly 2011 Model | Quarterly 2012 Model | Quarterly 2013 Model | Quarterly 2014 Model | Quarterly 2015 Model |
|---|---|---|---|---|---|---|
| **Complaint** | | | | | | |
| Yes | -1.684*** | -0.718 | -1.518*** | -2.589*** | -1.837*** | -1.443** |
| | (0.572) | (0.531) | (0.441) | (0.527) | (0.471) | (0.4857) |
| **Resolved** | | | | | | |
| Yes | 2.182*** | 1.006** | 1.473*** | 2.879*** | 2.024*** | 2.013*** |
| | (0.613) | (0.413) | (0.458) | (0.527) | (0.395) | (0.3957) |
| **Reason for contact** | | | | | | |
| *Water - related* | -0.348 | -0.609 | -0.594 | 0.966* | -0.136 | -0.286 |
| | (0.567) | (0.447) | (0.467) | (0.527) | (0.469) | (0.466) |
| *Waste - related* | 0.893 | 0.314 | 0.106 | 1.287** | 0.608 | -0.253 |
| | (0.687) | (0.492) | (0.5) | (0.582) | (0.535) | (0.514) |
| **Age** | | | | | | |
| *Under 35* | 0.884 | 0.681 | 0.314 | -0.0345 | 0.0309 | 0.0726 |
| | (0.661) | (0.486) | (0.518) | (0.599) | (0.505) | (0.537) |
| *45 - 54* | 2.558*** | 0.504 | 0.849 | -0.145 | 1.108* | -0.659 |
| | (0.794) | (0.535) | (0.573) | (0.609) | (0.584) | (0.486) |
| *55 - 64* | 2.277*** | 1.057* | 0.148 | 0.217 | 0.437 | 0.648 |
| | (0.86) | (0.594) | (0.532) | (0.617) | (0.56) | (0.564) |
| *65+* | 1.773** | 0.216 | 0.459 | 0.865 | 0.861* | 0.561 |
| | (0.869) | (0.596) | (0.585) | (0.818) | (0.519) | (0.569) |
| **Tenure** | | | | | | |
| *Owned* | -0.793 | -0.7 | -0.886** | -1.410*** | -0.538 | -0.247 |
| | (0.644) | (0.452) | (0.446) | (0.519) | (0.483) | (0.430) |
| Constant cut1 | -0.893 | -2.859*** | -2.312*** | -1.824** | -1.858*** | -2.072*** |
| | (0.808) | (0.688) | (0.638) | (0.752) | (0.646) | (0.597) |
| Constant cut2 | 0.242 | -2.285*** | -1.538** | -1.058 | -0.988 | -1.101 |
| | (0.793) | (0.638) | (0.606) | (0.72) | (0.61) | (0.565) |
| Constant cut3 | 0.848 | -1.151* | -0.587 | -0.307 | -0.128 | -0.362 |
| | (0.805) | (0.589) | (0.591) | (0.708) | (0.601) | (0.561) |
| Constant cut4 | 2.047** | 0.374 | 0.625 | 1.351* | 1.421** | 0.82 |
| | (0.838) | (0.581) | (0.595) | (0.724) | (0.618) | (0.568) |
| **Observations** | **99** | **139** | **123** | **118** | **139** | **143** |
| **Pseudo R2** | 0.2113 | 0.0891 | 0.1349 | 0.229 | 0.1506 | 0.1503 |
| **LR chi2(9)** | 49.74 | 28.67 | 46.24 | 66.75 | 54.72 | 55.41 |
| **Prob > chi2** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

*Source: Deloitte analysis*

### B.1.3 Predicted impacts on CSAT scores

The magnitude of the coefficients presented in Section B.1.2 needs to be understood in the context of the ordered logistic model. To estimate the magnitude of the impact of the individual variables on the CSAT scores it is possible to use the marginal effects of each variable to estimate the probabilities at each point of the continuous variables or at each category of the categorical variables.

For instance, based on the 2015 quarterly model in Table 18, the coefficient estimate for a contact that was categorised as a complaint, can be interpreted as follows. A complaint contact relative to a non-complaint contact would be expected to decrease by 1.4 the log of the odds of submitting a higher CSAT score level, all other variables being held constant. The estimated marginal effects for ordered logistic regressions can be interpreted as the change in probability of obtaining a particular CSAT score in response to a change in a given explanatory variable. The marginal effect also depends on the values taken by all the other covariates. Thus, it is common to evaluate marginal effects at the mean values of the covariates. For ordered logistic regressions, marginal effects need to be computed for each CSAT level separately.

Based on the estimated marginal effects, it is possible to estimate a series of predicted probabilities of each possible outcome (in this case a ranking of 1-5 for the CSAT score), which are, together, known as predictive margins.

Table 19 shows the predictive margins that a customer with specific characteristics would submit a “Very Satisfied” CSAT score. The rest of the explanatory variables are set at their respective mean values.

For example, a customer who owns (as opposed to rents) their accommodation has a 48% chance of submitting a CSAT score of 5 (or very satisfied), while a homeowner has a 52% probability. The probabilities associated with the contact attributes “Complaint” and “Resolved” indicate that there is a significant difference between customer contacts classified as a complaint or not and whether the matter has been resolved or not. If the reason for contacting the company was resolved at the time of the survey, the respondent has a 61% probability of submitting the highest CSAT score; the probability drops to 20% if the matter was not resolved.

**Table 19: Predictive margins: Quarterly TWUL Models based on Q4 2014-15 survey**

| Variables | Probability of Very satisfied | z-value | P > \|z\| |
|---|---|---|---|
| **Reason for contact** | | | |
| *Billing related* | 47% | 6.67 | 0.00 |
| *Waste related* | 56% | 9.49 | 0.00 |
| *Water related* | 45% | 8.3 | 0.00 |
| **Tenure (home ownership)** | | | |
| *Owned* | 48% | 10.79 | 0.00 |
| *Not owned* | 52% | 7.95 | 0.00 |
| **Complaint** | | | |
| *Yes* | 20% | 3.06 | 0.00 |
| *No* | 55% | 13.37 | 0.00 |
| **Resolved** | | | |
| *Yes* | 61% | 13.19 | 0.00 |
| *No* | 20% | 3.8 | 0.00 |

*Source: Deloitte analysis*

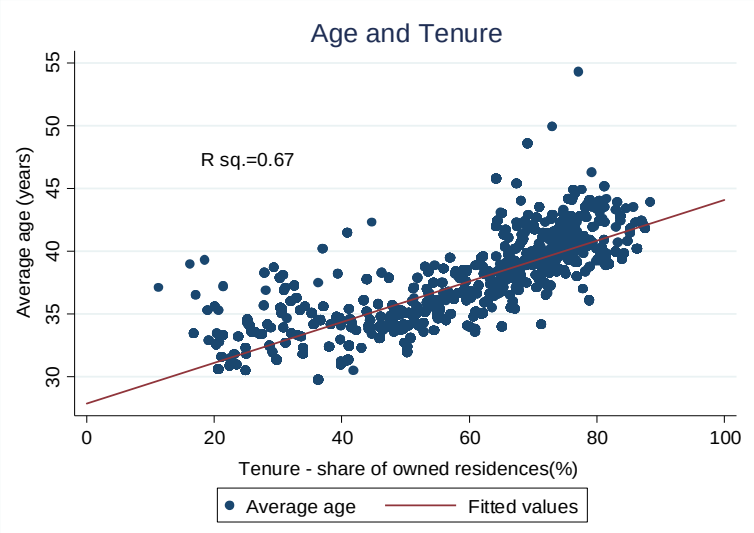
## B.2 Weekly TWUL Models

### B.2.1 Data characteristics

The weekly CSAT scores were analysed using ordered logistic regressions. Unlike the quarterly surveys (see Appendix B.1), the weekly surveys do not collect a number of respondents' socio-economic characteristics. Instead, the dataset was augmented with publicly available socio-economic data at the postcode district level, including average age, home ownership and qualifications.

As shown in Figure 6, average age and the share of homeowners within each postcode district are highly correlated. This implies that it is not possible to include these variables together in a regression. Therefore, a stepwise approach to the econometric analysis was used where each of the highly correlated variables were included one at a time in the model specifications presented in this study.

**Figure 6: Correlation among selected explanatory variables (Weekly TWUL Models)**




*Source: Deloitte analysis*

### B.2.2 Econometric results

The focus of the Weekly TWUL Models is to assess whether socio-economic characteristics of survey respondents have a statistically significant impact on the CSAT scores. In each model specification the control variables include contact attributes, such as the method of contact (written vs. telephone), the main reason for contacting the water company (billing, water, waste), whether the contact was categorised as a complaint, and the survey duration.

Table 20 presents the results for the different Weekly TWUL Model specifications. The results show that socio-economic characteristics tend not to be significant across models, while a number of contact attributes are consistently significant. In particular:

- **Resolution.** Respondents whose issue had been resolved at the time of the survey are associated with a higher CSAT score (Models 1-7 and Models 9-10).
- **Complaint.** Respondents whose contact was categorised as a complaint are associated with a lower satisfaction score (Models 1-10).
- **Survey duration.** Longer survey duration was associated with a higher CSAT score (Model 8), but the effect disappeared when the model controlled for "Contact resolved" (Model 2). In other words, the higher satisfaction score from surveys that took place over a longer period of time were likely to be impacted by the fact that the issue was more likely to have been resolved.

- **Reason for contact.** Contacts related to “Water” were associated with lower CSAT scores (Models 1-10), compared to those associated with “Billing”. There was no statistically significant difference between contacts related to “Waste” and “Billing” (Models 1-10). The results are statistically significant only for water related reasons for contacting the company and tend to result in lower scores, relative to billing related reasons.
- **Method of contact.** Contacts made in writing (as opposed to phone) are associated with lower satisfaction scores (Models 1-10).
- **Proximity to water treatment plant.** Customer proximity to a specific water treatment works was found not to be statistically significant (Models 1 and 10).
- **Socio-economic characteristics.** Only customer gender is statistically significant (male customers are associated with lower CSAT scores compared to female customers) in Models 9 and 10. Average age of the respondents was found to be statistically significant with a negative impact only in Model 8, where the resolution of the complaint was not taken into account. When resolution is included, average age loses its statistical significance (Models 3, 9 and 10). None of the other socio-economic characteristics tested are statistically significant.

## Table 20: Ordered logistic models (Weekly TWUL Models)

| Explanatory variables | Model Weekly 1 | Model Weekly 2 | Model Weekly 3 | Model Weekly 4 | Model Weekly 5 | Model Weekly 6 | Model Weekly 7 | Model Weekly 8 | Model Weekly 9 | Model Weekly 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Complaint** | | | | | | | | | | |
| Yes | -1.446*** | -1.447*** | -1.447*** | -1.445*** | -1.446*** | -1.448*** | -1.448*** | -1.770*** | -1.448*** | -1.448*** |
| | (0.0561) | (0.0561) | (0.0561) | (0.0561) | (0.0561) | (0.0561) | (0.0561) | (0.0547) | (0.0562) | (0.0562) |
| **Resolved** | | | | | | | | | | |
| Yes | 2.415*** | 2.414*** | 2.414*** | 2.415*** | 2.415*** | 2.412*** | 2.412*** | | 2.411*** | 2.411*** |
| | (0.0512) | (0.0512) | (0.0512) | (0.0512) | (0.0512) | (0.0512) | (0.0512) | | (0.0513) | (0.0513) |
| **Reason for contact** | | | | | | | | | | |
| *Waste related* | 0.0683 | 0.0682 | 0.0705 | 0.0609 | 0.0672 | 0.06 | 0.0646 | -0.218*** | 0.0787 | 0.0785 |
| | (0.0568) | (0.0568) | (0.0577) | (0.0586) | (0.0582) | (0.0573) | (0.0569) | (0.0549) | (0.0578) | (0.0578) |
| *Water related* | -0.211*** | -0.211*** | -0.209*** | -0.215*** | -0.211*** | -0.214*** | -0.213*** | -0.494*** | -0.204*** | -0.204*** |
| | (0.055) | (0.055) | (0.0558) | (0.0557) | (0.0557) | (0.0551) | (0.0551) | (0.053) | (0.0558) | (0.0558) |
| **Method of contact** | | | | | | | | | | |
| Writing | -0.287*** | -0.287*** | -0.287*** | -0.287*** | -0.287*** | -0.283*** | -0.284*** | -0.436*** | -0.282*** | -0.282*** |
| | (0.0628) | (0.0629) | (0.0629) | (0.0628) | (0.0629) | (0.0629) | (0.0629) | (0.0597) | (0.0629) | (0.0629) |
| **Presence of odour** | | | | | | | | | | |
| *Yes* | 0.119 | | | | | | | | | 0.124 |
| | (0.183) | | | | | | | | | (0.183) |
| **Survey duration** | | | | | | | | | | |
| Short | | -0.0267 | | | | | | -0.0913** | | |
| | | (0.0445) | | | | | | (0.0422) | | |
| **Average age** | | | -0.00147 | | | | | -0.0164*** | -0.00126 | -0.00107 |
| | | | (0.00663) | | | | | (0.00629) | (0.00663) | (0.00664) |
| **Tenure - share owned** | | | | 0.000686 | | | | | | |
| | | | | (0.00133) | | | | | | |
| **Tenure - share rented** | | | | | -0.000207 | | | | | |
| | | | | | (0.00245) | | | | | |
| **Highest qualification level** | | | | | | -0.00236 | | | | |
| | | | | | | (0.0021) | | | | |
| **No qualifications** | | | | | | | 0.00455 | | | |
| | | | | | | | (0.00468) | | | |
| **Gender** | | | | | | | | | | |
| *Male* | | | | | | | | | -0.0969** | -0.0974** |
| | | | | | | | | | (0.044) | (0.044) |
| Constant cut1 | -1.444*** | -1.462*** | -1.499*** | -1.410*** | -1.451*** | -1.535*** | -1.369*** | -3.579*** | -1.539*** | -1.530*** |
| | (0.0615) | (0.0671) | (0.25) | (0.0933) | (0.0833) | (0.101) | (0.1) | (0.236) | (0.251) | (0.251) |
| Constant cut2 | -0.864*** | -0.882*** | -0.920*** | -0.830*** | -0.871*** | -0.955*** | -0.789*** | -3.099*** | -0.959*** | -0.950*** |
| | (0.0587) | (0.0646) | (0.249) | (0.0915) | (0.0813) | (0.0989) | (0.0987) | (0.235) | (0.25) | (0.25) |
| Constant cut3 | -0.0946 | -0.113* | -0.15 | -0.0603 | -0.101 | -0.185* | -0.0192 | -2.504*** | -0.189 | -0.18 |
| | (0.0579) | (0.0638) | (0.249) | (0.0911) | (0.0807) | (0.0982) | (0.0983) | (0.234) | (0.25) | (0.25) |
| Constant cut4 | 1.179*** | 1.161*** | 1.124*** | 1.214*** | 1.173*** | 1.089*** | 1.255*** | -1.524*** | 1.085*** | 1.094*** |
| | (0.0605) | (0.0662) | (0.25) | (0.0928) | (0.0826) | (0.0997) | (0.1) | (0.233) | (0.25) | (0.251) |
| **Observations** | **9,146** | **9,146** | **9,146** | **9,146** | **9,146** | **9,146** | **9,146** | **9,146** | **9,146** | **9,146** |
| **Pseudo R2** | 0.1659 | 0.1657 | 0.1657 | 0.1657 | 0.1657 | 0.1657 | 0.1657 | 0.0573 | 0.1659 | 0.1659 |
| **LR chi2** | 3,724 | 3,718 | 3,718 | 3,718 | 3,718 | 3,719 | 3,719 | 1,285 | 3,723 | 3,723 |
| **Prob > chi2** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

*Source: Deloitte analysis*

The Brant test was run for all the Models 1 to 10 to test the proportional odds assumption. The results consistently show that the variables representing customer attributes pass the test while the contact characteristics variables violate the assumption. Under these circumstances, the alternative approach would be to estimate generalized ordered logit models, where each customer satisfaction level is associated with the contact and customer attributes separately.

For each of the TWUL Weekly model specifications presented above, separate generalized ordered logistic regressions were estimated. The generalised ordered regressions yielded similar coefficient estimates for all the variables in terms of both magnitude and significance to the ordered logit regressions. As a result, the weekly results showed in Table 20 could be an acceptable approximation of the variables' impacts.

### B.2.3 Predicted impacts on CSAT scores

As set out in Section B.1.3, the interpretation of the regression results is in terms of the predicted probabilities for individual customers submitting a given CSAT score. The predicted impacts for selected variables from the Weekly TWUL Models are shown in Table 21.

Table 21 shows the estimated probabilities for the CSAT scores being Very satisfied across the categories of several explanatory variables. For example, the probability that the CSAT score takes the value 5 (representing the option very satisfied) is 56% given that the main reason for contact is related to Billing and the rest of the variables are set to their mean values. Furthermore, the probability of a CSAT score of 5 is very high (70%) if the matter for which the responded contacted the company has been resolved.

**Table 21: Predictive margins: Selected variable in TWUL Weekly models**

| Variables | Probability of Very satisfied | z-value | P > \|z\| |
|---|---|---|---|
| **Reason for contact** | | | |
| *Billing related* | 56% | 82.62 | 0.00 |
| *Waste related* | 59% | 86.52 | 0.00 |
| *Water related* | 54% | 74.71 | 0.00 |
| **Survey duration*** | | | |
| *1 week (short)* | 58% | 80.41 | 0.00 |
| *2 weeks (long)* | 56% | 89.54 | 0.00 |
| **Complaint** | | | |
| *Yes* | 32% | 31.47 | 0.00 |
| *No* | 60% | 129.28 | 0.00 |
| **Resolved** | | | |
| *Yes* | 70% | 128.39 | 0.00 |
| *No* | 19% | 27.96 | 0.00 |

**Estimated without controlling for resolution of complaint.*

*Source: Deloitte analysis*

## B.3 Panel Models

### B.3.1 Data characteristics

The overall CSAT scores for each company over time have been analysed within a panel that stacks the scores by company over time. Using this structure, it is possible to assess the impact that regional and company differences may have on CSAT scores.

The Panel Models are intended to disentangle the regional and the company attributes and their relative effects on the overall CSAT scores. To do so, the following inputs were used:

- Average socio-demographic profiles for each company area using the 2011 Census data, ONS estimates of net weekly income (averaged across individual companies) and data from Ofwat on average hourly wages (from 2012/13); and
- Company attributes including WoC and WaSC, company operating expenditure per customer, location, geography, population turnover and size.

Furthermore, since most of the explanatory variables are categorical and thus time-invariant, it is not possible to estimate Panel Models with fixed effects as they would be omitted in the regressions. On the other hand, it is possible to estimate panels with random effects, where the idiosyncratic and unobserved company effects are taken into account. Further details on the choice between random effects and fixed effects models is provided in Appendix A.3.2.

Two sets of models were estimated to reflect the nature of the explanatory variables:

- Categorical-only models. These models (Set 1 and 2 below) use only time-invariant explanatory variables, including for income (as set out in A.2.3).
- Models with income levels. These models (Set 3 below) included the levels of average income, by company area, as explanatory variables (although these were only available annually, rather than quarterly, with the most recent data covering 2012/13).

Finally, the number of regressors included in individual models below tends to be limited due to a high degree of correlation among them (a problem known as 'multicollinearity'). If a large number of explanatory variables was included, this would either render their coefficients insignificant or they would be omitted in the regressions. For illustration, Table 22 shows the correlations between a set of company attributes.[27] The estimated correlation coefficients show that company size tends to be highly correlated with company type and company location.

**Table 22: Illustrative correlations between company attributes**

| | Company type | Company size | Company location | Turnover |
|---|---|---|---|---|
| **Company type** | 1 | | | |
| **Company size** | 0.80 | 1 | | |
| **Company location** | 0.04 | -0.72 | 1 | |
| **Turnover** | -0.65 | -0.35 | -0.18 | 1 |

*Source: Deloitte analysis*

[27] The levels of the categorical variables are converted into 0 and 1 in order to compute the correlation coefficients. For instance, company size category "H" and company type "wasc" are encoded as 1. The large and positive correlation coefficient (0.8) between these two variables implies that WaSCs are highly correlated with the companies being large.

### B.3.2 Econometric results

Since most of the explanatory variables in the Panel Models are categorical (as opposed to continuous), except for average real wages and total operating expenditure by customer across companies, the interpretation of the coefficients needs to account for the two distinct categories that these variables take (typically “high” or “low”). As a result, the Panel Models can identify characteristics of companies and of their populations that, if they are higher or lower than average, have a statistically significant impact on CSAT scores. However, they cannot capture the variation within the discrete categories.

As shown in Table 23 and Table 24, the Panel Models suggest that:

**Company attributes.** Customers of WoCs are associated with a higher CSAT score compared to WaSCs, although this is not statistically significant in model 6. Other company attributes such as total operating expenditure by customer size, location (North vs South) and population turnover are not statistically significant in the wide majority of models estimated.

- **Customer socio-economic characteristics** tend not to be statistically significant. Income measures tend to be significant in some, but not all, model specifications. In particular:
  - Companies with customers who have higher than average real hourly wage levels or higher than average weekly income are associated with lower CSAT scores (Models 11, 12).
- **Income measures.** It is not possible to disentangle the individual impacts of income, disability and home ownership, since these variables are highly correlated with each other. There are four key models that indicate that the socio-economic characteristics of the population served by a given water company are highly significant for the expected CSAT scores. In particular, companies with a population that is “better off” than the average population in terms of these characteristics tend to report higher scores. These include:
  - Tenure: lower than average home ownership is associated with higher scores (Model 13);
  - Net weekly income[28]: lower than average income is associated with higher scores (Models 1 - 10 and 12); and
  - Real hourly wage: lower than average wage is associated with higher scores (Model 11).

[28] The detailed description of the data is provided in Appendix A.1.

## Table 23: Panel Models Set 1: categorical models

| Explanatory variables | Panel Model 1 | Panel Model 2 | Panel Model 3 | Panel Model 4 | Panel Model 5 | Panel Model 6 | Panel Model 7 | Panel Model 8 | Panel Model 9 | Panel Model 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Time trend** | 0.00394* | 0.00394* | 0.00394* | 0.00394* | 0.00394* | 0.00394* | 0.00394* | 0.00394* | 0.00415* | 0.00394* |
| | (0.00223) | (0.00223) | (0.00223) | (0.00223) | (0.00223) | (0.00223) | (0.00223) | (0.00223) | (0.00225) | (0.00223) |
| **Company type** | | | | | | | | | | |
| *WOC* | 0.167*** | 0.168*** | 0.177*** | 0.147** | 0.163*** | 0.162* | 0.167*** | 0.171*** | | 0.149** |
| | (0.0527) | (0.0566) | (0.0547) | (0.0607) | (0.0519) | (0.0922) | (0.0507) | (0.0616) | | (0.0644) |
| **Weekly net income** | | | | | | | | | | |
| *L* | 0.218*** | 0.225*** | 0.200*** | 0.217*** | 0.276*** | 0.218*** | 0.220*** | 0.211*** | 0.154* | 0.253*** |
| | (0.0665) | (0.0537) | (0.0632) | (0.0667) | (0.0656) | (0.0645) | (0.0593) | (0.0801) | (0.0791) | (0.0855) |
| **Age** | | | | | | | | | | |
| *Younger* | -0.0072 | | | | | | | | | |
| | (0.0944) | | | | | | | | | |
| **Disability** | | | | | | | | | | |
| *Lower* | | 0.0102 | | | | | | | | |
| | | (0.0603) | | | | | | | | |
| **Ethnicity** | | | | | | | | | | |
| *Lower* | | | -0.0741 | | | | | | | |
| | | | (0.0616) | | | | | | | |
| **Tenure** | | | | | | | | | | |
| *Lower than average homeownership share* | | | | 0.111* | | | | | | |
| | | | | (0.0571) | | | | | | |
| **Status** | | | | | | | | | | |
| *DE* | | | | | -0.0796 | | | | | |
| | | | | | (0.0577) | | | | | |
| **Company size** | | | | | | | | | | |
| *L* | | | | | | 0.00764 | | | 0.101 | |
| | | | | | | (0.0977) | | | (0.0684) | |
| **Location** | | | | | | | | | | |
| *Urban* | | | | | | | -0.00196 | | | |
| | | | | | | | (0.0727) | | | |
| **Geography** | | | | | | | | | | |
| *S* | | | | | | | | -0.0191 | | |
| | | | | | | | | (0.0886) | | |
| **Company expnditures** | | | | | | | | | | |
| *(by customer)* | | | | | | | | | -0.000563 | |
| | | | | | | | | | (0.00302) | |
| **Population turnover** | | | | | | | | | | -0.0549 |
| *L* | | | | | | | | | | (0.0884) |
| **Constant** | 4.072*** | 4.056*** | 4.096*** | 4.023*** | 4.068*** | 4.065*** | 4.067*** | 4.082*** | 4.141*** | 4.078*** |
| | (0.125) | (0.0868) | (0.0797) | (0.0765) | (0.0681) | (0.0675) | (0.0841) | (0.0974) | (0.133) | (0.0719) |
| **Observations** | **408** | **408** | **408** | **408** | **408** | **408** | **408** | **408** | **404** | **408** |
| Number of companies included | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 |
| **Wald chi2** | 37.49 | 35.98 | 45.77 | 28.8 | 32.99 | 36.48 | 36.41 | 37.28 | 8.65 | 27.1 |
| **Prob > chi2** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.07 | 0.00 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

*Source: Deloitte analysis*

*Note: Clustered robust standard errors were used in this set of results.*

**Table 24: Panel Models Set 2: categorical models**

| Explanatory variables | Panel Model 11 | Panel Model 12 | Panel Model 13 | Panel Model 14 |
|---|---|---|---|---|
| **Time trend** | 0.00341 | 0.00394* | 0.00341 | 0.00341 |
| | (0.00217) | (0.00223) | (0.00217) | (0.00217) |
| **Company type** | | | | |
| *WOC* | 0.162*** | 0.167*** | 0.0878 | 0.0975 |
| | (0.0503) | (0.0533) | (0.0737) | (0.0713) |
| **Real wages** | | | | |
| *LW* | 0.216*** | | | |
| | (0.0585) | | | |
| **Weekly net income** | | | | |
| *L* | | 0.220*** | | |
| | | (0.0588) | | |
| **Tenure** | | | | |
| *Lower than average homeownership share* | | | 0.0971 | |
| | | | (0.0771) | |
| **Disability** | | | | |
| *Lower* | | | | -0.087 |
| | | | | (0.0734) |
| **Constant** | 4.075*** | 4.065*** | 4.209*** | 4.291*** |
| | (0.0693) | (0.0688) | (0.0649) | (0.0705) |
| **Observations** | **432** | **408** | **432** | **432** |
| Number of companies included | 18 | 17 | 18 | 18 |
| **Wald chi2** | 39.82 | 35.88 | 7.76 | 6.14 |
| **Prob > chi2** | 0.00 | 0.00 | 0.05 | 0.10 |
| Robust standard errors in parentheses | | | | |
| *** p<0.01, ** p<0.05, * p<0.1 | | | | |

*Source: Deloitte analysis*
*Note: Clustered robust standard errors were used in this set of results.*

In addition to the results presented in Set 1 and Set 2 above, the relationship between CSAT scores and real wages has been tested using a set of models with income levels (as opposed to categorical variables). In these results, shown as Models 15 to 24 in Table 25, a higher real wage remains associated with lower CSAT scores across a number of specifications.

## Table 25: Panel Models Set 3: models with income levels

| Explanatory variables | Panel Model 15 | Panel Model 16 | Panel Model 17 | Panel Model 18 | Panel Model 19 | Panel Model 20 | Panel Model 21 | Panel Model 22 | Panel Model 23 | Panel Model 24 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Time trend** | 0.00054 | 0.00055 | 0.000798 | 0.000375 | 0.000464 | 0.000611 | 0.000509 | 0.000703 | 0.00133 | 0.000508 |
| | (0.00188) | -0.00187 | (0.0018) | (0.00196) | (0.00179) | (0.00186) | (0.00188) | (0.00189) | (0.00187) | (0.00187) |
| **Company type** | | | | | | | | | | |
| *WOC* | 0.135** | 0.134** | 0.144** | 0.126** | 0.134** | 0.114 | 0.134** | 0.147** | | 0.137** |
| | (0.0586) | (0.0613) | (0.0586) | (0.0572) | (0.057) | (0.0881) | (0.0586) | (0.0728) | | (0.0671) |
| **Real wages** | -0.0573*** | -0.0571*** | -0.0522** | -0.0606*** | -0.0588*** | -0.0559*** | -0.0579*** | -0.0541** | -0.0482** | -0.0580** |
| | (0.0202) | (0.019) | (0.0208) | (0.0214) | (0.0203) | (0.0201) | (0.0193) | (0.0221) | (0.0238) | (0.024) |
| **Age** | | | | | | | | | | |
| *Younger* | -0.0142 | | | | | | | | | |
| | (0.08) | | | | | | | | | |
| **Disability** | | | | | | | | | | |
| *Lower* | | -0.0087 | | | | | | | | |
| | | (0.0592) | | | | | | | | |
| **Ethnicity** | | | | | | | | | | |
| *Lower* | | | -0.0747 | | | | | | | |
| | | | (0.0576) | | | | | | | |
| **Tenure** | | | | | | | | | | |
| *Lower than average homeownership share* | | | | 0.0983* | | | | | | |
| | | | | (0.0557) | | | | | | |
| **Status** | | | | | | | | | | |
| *DE* | | | | | -0.00305 | | | | | |
| | | | | | (0.061) | | | | | |
| **Company size** | | | | | | | | | | |
| *L* | | | | | | 0.0335 | | | 0.103 | |
| | | | | | | (0.0891) | | | (0.0646) | |
| **Location** | | | | | | | | | | |
| *Urban* | | | | | | | 0.00117 | | | |
| | | | | | | | (0.0681) | | | |
| **Geography** | | | | | | | | | | |
| *S* | | | | | | | | -0.0396 | | |
| | | | | | | | | (0.0798) | | |
| **Company expnditures** | | | | | | | | | | |
| *(by customer)* | | | | | | | | | 0.00121 | |
| | | | | | | | | | (0.00295) | |
| **Population turnover** | | | | | | | | | | |
| *L* | | | | | | | | | | 0.00569 |
| | | | | | | | | | | (0.0843) |
| **Constant** | 5.207*** | 5.198*** | 5.127*** | 5.213*** | 5.224*** | 5.164*** | 5.207*** | 5.163*** | 4.994*** | 5.204*** |
| | (0.31) | (0.294) | (0.325) | (0.347) | (0.329) | (0.321) | (0.315) | (0.339) | (0.349) | (0.427) |
| **Observations** | **432** | **432** | **432** | **432** | **432** | **432** | **432** | **432** | **428** | **432** |
| Number of companies included | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 |
| **Wald chi2** | 14.08 | 14.29 | 21.44 | 20.9 | 14.21 | 14.77 | 13.59 | 13.92 | 8.24 | 14.56 |
| **Prob > chi2** | 0.01 | 0.01 | 0.00 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 | 0.08 | 0.01 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

*Source: Deloitte analysis*

*Note: Clustered robust standard errors were used in this set of results.*

# Appendix C
# Normalisation methodology

This Appendix sets out a potential normalisation methodology, based on literature review, for customer satisfaction scores. This methodology, unlike the alternatives set out in the main body of the report, does not rely on the econometric analysis and is provided for reference given its position in the academic literature.

## C.1 Worked example of the methodology

As set out in Section 3, a normalisation approach to CSAT scores could help address the potential differences in customer expectations in different regions by standardising the components of the satisfaction score (related to billing, water and waste).

Since customers' expectations are not observable, the approach indirectly uses customers' satisfaction with other services from the same company as a benchmark when evaluating customers' satisfaction with the service of interest. For example, the "raw" satisfaction scores on water and waste can be used as a benchmark when assessing whether the "raw" satisfaction score on billing is high or low for a particular company.

The limitation of this normalisation methodology is that it implicitly removes any company-specific factors that impact all services equally. As a result, the methodology fails to distinguish whether the relatively lower scores for a company are driven by the underlying customers' expectations, or whether they are driven by a relatively lower quality of the services provided.

Table 26 illustrates the normalisation process for a hypothetical set of scores. In this example, a regional organisation A receives customer satisfaction scores for three service lines, 1, 2, and 3. The weighted-average raw score for organisation A is calculated to be 13.75. The weights chosen for the worked example are consistent with the weights applied by the CSAT methodology to the WaSC data: 50% billing / 25% water operations / 25% waste water operation. In order to account for potential differences across A's customers, the raw scores for each service line is divided by A's overall score. This yields a value close to 1 for each service line.

**Table 26: Normalisation methodology: worked example**

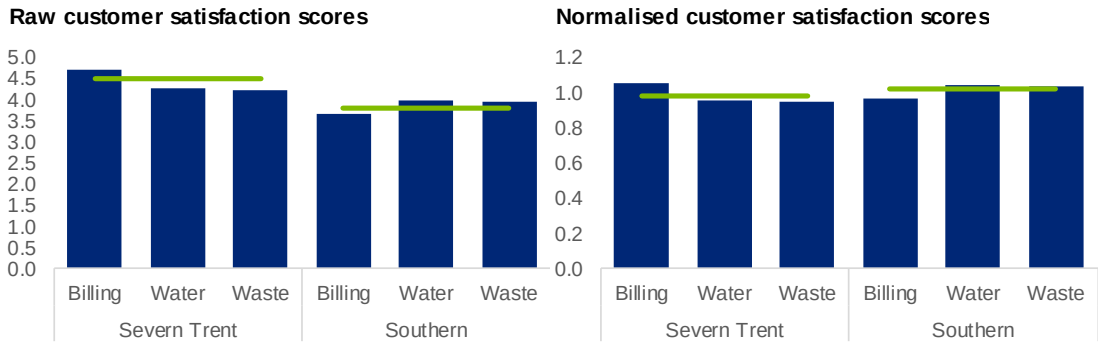
| Category | Score for Service 1 | Score for Service 2 | Score for Service 3 | Average score | Detail: calculation of average score |
|---|---|---|---|---|---|
| "Raw" scores for **A** | 10 | 20 | 15 | 13.75 | Weighted average score (50% for Service 1, 25% for Service 2 and 25% for Service 3) |
| Normalised scores for **A** | 10/13.75 = **0.72** | 20/13.75 = **1.45** | 15/13.75 = **1.09** | 1.09 | Unweighted Average of normalised scores |

*Source: Deloitte analysis*

A value above 1 would indicate that A's customers are relatively more satisfied with that particular service than the other services. The normalised scores for each service line can be directly compared across regional organisations (organisation A's service line 1 vs. organisation B's service line 1).

Applying this approach to water companies, Figure 7 shows the CSAT scores for two companies. The left hand side panel shows the methodology for calculating the CSAT scores using the "raw" scores, while the panel on the right shows the normalised approach.

**Figure 7: Normalisation of customer satisfaction scores: illustration for two companies**




*Source: Deloitte analysis based on normalisation methodology from Trompet et al. (2013)*
*Note: The green line represents the overall CSAT score, a weighted average of the Billing, Waste and Water scores.*

Using raw CSAT scores, in the left hand-side panel, Severn Trent Water is found to have more satisfied customers than Southern Water. However, if the assumption on different baseline service expectations is valid (and this is not observable), then this comparison fails to take into account the fact that perhaps Severn Trent Water's customers have a propensity to give a higher score, regardless of the quality of service provided.

By contrast, normalised customer satisfaction scores, shown in the right hand-side panel in Figure 7 indicate that Southern Water actually performed relatively better than Severn Trent Water, with respect to their respective customer expectations.

## References

Johnson, M., Herrmann, A., Gustafson A. (2002): Comparing customer satisfaction across industries and countries. Journal of Economic Psychology, 23, 749 – 769.

Trompet, M., Parasram, R, Anderson, R.J, *Benchmarking Disaggregate customer satisfaction scores between bus operators in different cities and countries*. Transportation Research Board, 2013

The Transit Cooperative Research Program (TCRP). *Report 141: A Methodology for Performance Measurement and Peer Comparison in the Public Transportation Industry*. Transportation Research Board of the National Academies, Washington D.C. 2010.

UN Guidelines: Chapter 3: UN Department of Economic and Social Affairs – Statistics Division. Designing Household Survey Samples: Practical Guidelines, 2005

Eurostat Methodologies and Working Papers: Handbook on precision requirements and variance estimation for ESS household surveys, 2013.
http://ec.europa.eu/eurostat/documents/3859598/5927001/KS-RA-13-029-EN.PDF