South Staffs Water

**South Staffs Water PR14 Stated Preference Survey:
Report submitted to South Staffs Water by ICS and eftec
June 2013**

A review

Ken Willis,
Newcastle University

17th June 2013

Key findings

1. The study is an excellent piece of research.
2. It adheres to the standard process for conducting these types of studies: from focus groups, experimental design, pilot surveys, through to econometric models, and validity checks.
3. The study sensibly assembled the attributes into three groups, with 3 or 4 attributes in each group, which allows customers to easily trade-off attributes against each other.
4. The sample is representative of both domestic and business customers, and the sample size is sufficient to produce statistically significant results.
5. An appropriate choice model was developed: a state of the art mixed logit model, to account for heterogeneity in customer preferences, was rightly used to estimate preferences and utility for attributes.
6. The WTP amounts are accurate and robust.

Introduction

The South Staffs Water Stated Preference Survey undertaken by ICS and eftec is an exceptionally good piece of research. The research is comprehensive and meticulous in its approach; and it is "state of the art" in the application of stated preference methods.

The research fulfils the terms of reference for the project which is to:

1. Estimate the value to customers, in money terms, of changes in water service levels;
2. Ensure the values are appropriate for use in cost-benefit analysis (CBA);
3. Build upon the work and outputs of recent UKWIR studies on the application of willingness-to-pay (WTP) studies and CBA.

Methodology

The report provides a very articulate and eloquent description of the methodology involved in a stated preference study, which should easy for non-economists to understand.

There are a number of issues in applying stated preference methods such as contingent valuation (CV) and choice experiments (CEs). Amongst these are

1. accounting for non-linear effects, such as diminishing marginal utility, and gain-loss asymmetry in utility, and between willingness-to-pay (WTP) improvements and willingness-to-accept (WTA) bill reductions for deteriorations in service;

2. accounting for income effects and substitution effects when services are valued separately (rather than simultaneously), and then summed to derive a total value: the individual valuation and summation problem.

It is important address both of these issues in a valuation study; and the study by ICS & eftec does so admirably.

Stated Preference Research

The report on the study provides a good overview of the tasks involved in a stated preference research, namely

- Identify the service measures that South Staffs Water (SSW) customers consider important in water supply and quality, and the environment; and identify realistic changes in levels for these service measures (SMs);
- Undertake qualitative research (focus groups, and cognitive interviews) to test customers' understanding of SMs descriptions, the questionnaire, and illustrative material presented to them;
- design an experiment to allow the value of each attribute or SM to be identified;
- undertake a pilot survey to test that the questionnaire and choice experiments are working;
- elicit customers' preferences from the pilot survey data using stated choice experiments (CEs) that allow customers to trade-off SM levels, and to choose between alternative bundles of attribute levels;
- develop a sampling strategy, and survey a representative sample of domestic and business customers;
- specify the econometric models to analyse choice, estimate utility, and willingness-to-pay (WTP) values for changes in the level of each attribute;
- analyse the data, and estimate WTP for improvements in attribute levels, and willingness-to-accept (WTA) compensation (i.e. a bill reduction) for any deterioration in attribute levels;
- evaluate the validity of the responses.

Qualitative Research

The approach by ICS & eftec to the qualitative research, testing customer preferences and demands for service improvements through a combination of service descriptions, and illustrative material (photographs, diagrams, and statistical information) in the focus groups, is commendable. The qualitative research survey instruments were carefully designed, and expertly administered.

The focus groups with domestic customers, undertaken by Opinion Leader, were adequate to identify the service measures customers considered important, and ensure the service measures descriptions were understood by respondents and that sufficient information was available to respondents. The focus groups rightly comprised a mix of customers in each group, by gender, socio-economic group (SEG), ethnicity, and ability to pay. Although the report mentions 6 focus groups, Table 3.1 only documents 4 focus group types. Did some types include more than one focus group?

The draft questionnaire was tested by 10 cognitive interviews. This is an admirable approach. Cognitive testing ensures that the questionnaire, attribute descriptions and material, were fully understood by customers, and also that the choice experiments (CEs) could be answered by respondents, prior to the pilot survey.

The pilot survey of 100 domestic customers was well conducted. The pilot study revealed that customers could easily understand the levels of service, and their descriptions, the levels of service were realistic to customers; and that customers were able to make comparisons between the choices presented on the choice cards.

The pilot survey identified a few small areas where wording and presentation in the questionnaire was capable of improvement, which were addressed prior to the main survey. The pilot survey also revealed the need to revise the range of bill amounts changes, and the need to shorten the length of the questionnaire.

Choice Experiment Format
The study sensibly assembled the attributes into three groups, with three or four attributes in each group. The small number of attributes in each group allows customers to easily trade-off attributes against each other. The number of attributes which customers can simultaneously trade-off is limited to around 5 or 6, before respondents adopt some heuristic to ease the choice task. It is important that customers simultaneously trade-off all attributes, rather than adopt some heuristic which is likely to bias choice and result in the choice models deriving biased estimates of customers' "true" preferences.

The study asked customers to choose between two alternative attribute bundles, plus the status quo. ICS & eftec provide a reasoned justification for the inclusion of the status quo situation on choice cards. The vast majority of participants in the focus groups thought that the current level of service was acceptable, and that an additional charge to the bill for an improvement in this service was not seen as necessary. This supports the need to include the status quo on each choice card, as a valid choice option for customers. It also avoids hypothetical bias in respondents being forced to choose one of two hypothetical alternatives if the status quo was omitted.

The survey specified price changes in percentage terms, but for households this was expressed in the questionnaire as monetary amounts based on the customer's current bill level. This more closely reflects the format (absolute change in amount) by which households' judge price changes, as well as in percentage terms.

Experimental design
The report outlines the principles involved in the experimental design for the study, but does not go into much detail. A D-optimal design was rightly adopted. But it would be useful to specify here the number of attributes in each of the choice experiment, and what the attributes are. It is only later that the reader discovers in Table 4.12 that the "drinking water quality" choice experiment (CE) contained 4 attributes, the "water availability and environment" CE 3 attributes, and the "reliability of water supply" CE 4 attributes.

A fractional factorial design was probably adopted, but how many choice alternatives did this generate for each of the 3 CEs? And how many choice cards, each with two hypothetical alternatives plus the status quo, did each respondent answer in each of the three CEs?

Sample

The main survey sample of 500 randomly selected domestic customers across SSW area, and 300 business customers, is sufficient to ensure that the sample is representative of SSW customers, and that the results are statistically significant.

The representativeness of the sample of 500 domestic customers was set by gender, age, and socio-economic group (SEG). The sample closely reflects the SSW region population age profile. SEGs are probably more important in that this grouping in more closely related to income and ability to pay. The sample was broadly representative of the regional SEG profile, although it over-represented C1/C2 groups, and under-represented D social group.

The business sample of 300 businesses provides a good representation of the industrial structure of the SSW region. The sample also closely matches the distribution of actual bill amounts paid by businesses.

So SSW can be confident that the results derived from the sample data will be representative of SSW domestic and business customers.

Consumer Characteristics

The report neatly documents the characteristics of the domestic customer samples in terms of employment, education, and income. This is helpful in envisaging the profile of the sample of domestic customers. Business consumer characteristics are portrayed in terms industrial structure, organisation size, and bill amounts.

The study also documents customers' reported experience of water service failures, within the past year, e.g. 10% of domestic respondents reported experiencing discoloured water with the last year, and 17% reported that they had experienced low water pressure within the last year. Clearly customer perceptions about service failures, such as discoloured tap water, and taste and smell of tap water, are much higher than the actual incidence of service failures as recorded by SSW; and used in Table 3.2: Service Levels, as the "base case" or current situation, which forms part of the experimental design for the study.

It is not uncommon in stated preference studies for actual and perceived chances and occurrences to differ. This is not a problem for the WTP estimates where customers' WTP is driven by perceived experiences and preferences. Moreover not all customers report every service failure. So customers' perceptions of service failures, and customers' attitudes (reflected in customers' priorities for improvements), are likely to be much more instrumental in determining WTP than recorded service failures.

Choice models

Domestic customers

The choice models adopted in the analysis are appropriate: a conditional logit (CL) model; an error component (EC) model to test for variance between choice alternatives; and a mixed logit (MXL) model analysis which allows for heterogeneity in preferences between customers.

The MXL model fitted the data best, so only these models are reported. The MXL models have high pseudo-R^2 values across all CEs, indicating a good fit to the data. The non-linear models fit the data marginally better than the linear models for all three CEs: "drinking water quality" (DWQ), "water availability and the environment" (WAE), and "reliability of water supply" (RWS).

However, specifications of the MXL models are not explained: that is, what attributes were given random coefficients and what distributions were used? The MXL models simply report the main coefficient for each of the service measures. Where are the estimates of the standard deviations of the random coefficients? They are not in Tables 4.13 to 4.18 along with the other estimates. So it is difficult to ascertain the extent of heterogeneity amongst customer preferences for different water service measures.

All of the coefficients for the service measure attributes have the right sign, and all are statistically significant, in all the three CEs, except for "hard water +1 level" in the DWQ block. Both these conditions (right sign and statistical significance) are necessary for accurate and reliable estimates of WTP for each attribute.

The coefficient for the water bill (deterioration) is not significant in the DWQ CE, and is only marginally significant in the RWS CE. However, the water bill (improvement) coefficients are highly statistically significant across all three CE, which is important in deriving accurate and reliable estimates of WTP for service measure improvements.

Business customers
Much the same comments can be made about the non-domestic customer choice experiments. All three MXL models provide an excellent fit to the data, with high pseudo-R^2 values. All the service measure coefficients have the right sign and are statistically significant, except for hard water +1 level, and hard water +2 level. This is somewhat anomalous, since the service measure hardness of tap water is judged by business customers as having the greatest priority for improvement (see Figure 5.1).

Again, the specifications of the MXL models are not explained: what attributes were given random coefficients and what distributions were used? Where are the estimates of the standard deviations of the random coefficients? Hence, as with the domestic MXL model results, it is difficult to judge the extent of heterogeneity amongst business customer preferences for different water service measures.

The coefficient for the water bill (deterioration) is not significant in the DWQ or in the RWS choice experiment. Hence, WTA values for deteriorations in service measures for attributes in these two blocks cannot be derived from these non-linear models. As the report points out, if WTA values are required then they should be derived from the linear MXL model for the DWQ and the RWS attributes.

WTP values
Domestic customers
The study rightly includes a contingent valuation (CV) question to ensure that the individual valuation of the three CE blocks of attributes does not exceed the combined WTP of customers for all three blocks of service measures.

The CV question provides strong evidence of that the summation of the individual CE WTP block values does indeed over-estimate total WTP. Moreover, 28% of customers also indicated that they would revise their WTP amount to SSW conditional on having to pay an increased sewerage bill to Severn Trent Water for improvements to sewerage services. Thus ICS & eftec are correct to use the CV WTP value as the "true" WTP value. And to "scale" the WTP amounts derived from the CE model, for each service measures, in each of the three CE blocks accordingly.

5

The study estimates a WTP value of £9.80 increase in domestic customer water bills for an improvement from the status quo to level +2. This increase is comparable to some studies for other water companies for improvements across this range of attributes.

This amount also seems reasonable in the light information from the questionnaire on attitudes to current bill levels (Table 4.10), and also customers' experience of water service failures (Table 4.9). The study also found that 77% of customers wanted future bills to remain the same and service levels unchanged (Table 4.11), whilst only 23% wanted bills to increase by a small amount of services to improve.

Business customers
The non-domestic survey also rightly included a contingent valuation (CV) question to ensure that the individual valuation of the three CE blocks of attributes did not exceed the combined WTP of business customers for all three blocks of service measures.

Again, the CV question provides strong evidence that the summation of the individual CE WTP block values would indeed over-estimate total WTP. ICS are correct in using the CV result to scale the CE WTP values in the individual blocks of attributes.

Conclusion
SSW can be assured that the WTP values derived by ICS & eftec for the service measure improvements are accurate, reliable, and robust. SSW can confidently use these WTP values in their CBA appraisal of investment programs and projects.

However, the WTP amounts derived from the CEs and CV analyses, are the maximum consumer surplus amounts for improvements to level +2. These are the total benefit customers would derive from the improvements: not the actual bill amount required to undertake the investment necessary to deliver the improvements. The cost of the improvements will usually be less than the total benefit consumers derive from the good.

Validity checks
The validity of the results can be judged in terms of theoretical validity: the goodness-of-fit of the models, and the signs and statistical significance of the coefficients. This shows that the models are theoretically valid, and therefore reliable and robust.

CE results also can be assessed in terms of content validity (of the questionnaire survey). The study is valid in terms of content validity: the respondent feedback presented in the study indicates the vast majority of domestic and business respondents had no difficulty in answering the choice tasks in the questionnaire. Interviewer feedback also confirmed the validity of the study: interviewers thought only 9% of domestic customers had limited understanding of the choice tasks, and that 93% of respondents gave serious consideration to the questions. So DCWW can be confident of the validity of the survey results.

Conclusions
The stated preference study by ICS & eftec for SSW is an excellent, commendable, and professional piece of research. The study is "state of the art" and conforms to best practice. The analysis is meticulous and detailed, and provides accurate and reliable information about customers' preferences. It provides a wealth of information on customers' WTP values which can be used in a cost-benefit analysis of investment projects to improve water supply and water quality to SSW customers.