



Research Paper 35 | 2015

USING DISCRETE CHOICE EXPERIMENTS TO REGULATE THE PROVISION OF WATER SERVICES: DO STATUS QUO CHOICES REFLECT PREFERENCES?

B. LANZ AND A. PROVINS

Using discrete choice experiments to regulate the provision of water services: Do status quo choices reflect preferences?*

Bruno Lanz [†]

Allan Provins[‡]

This version: February 2015

Abstract

Discrete choice experiments (DCE) are increasingly used to quantify the demand for improvements to services provided by regulated utility companies and inform price controls. This form of preference elicitation, however, often reveals a high frequency of status quo (SQ) choices. This may signal an unwillingness of respondents to evaluate the proposed trade-offs in service levels, questioning the welfare theoretic interpretation of observed choices and the validity of the approach for regulatory purposes. Using the methodology for DCE in the regulation of water and sewerage services in England and Wales, our paper contributes to the understanding of SQ choices in several novel dimensions. First, we control for the perception of the SQ and the importance of attributes in day-to-day activities. Second, we use a split sample design to vary both the description of the SQ and the survey administration mode (online vs. in-person). Third, the service attributes can both improve or deteriorate, so that the SQ is not necessarily the least-cost option. Fourth, we examine SQ choices in individual choice tasks and across all tasks so as to identify the determinants of serial SQ choices. Our results suggest that *individual* SQ choices mostly reflect preferences and thus represent important information for the regulator. However, *serial* SQ choices are mainly driven by cognitive and/or contextual factors, and these responses should be analysed as part of standard validity tests.

Keywords: Cost-benefit analysis; Regulated utilities; Economic valuation; Discrete choice experiments; Individual decision making; Status quo effect.

JEL classification: C25; L43; L95; Q25; Q51; Q58

*We thank Ian Bateman, Ali Chalak, Sergio Colombo, Scott Reid, Ken Willis, and participants at the EAERE conference for helpful comments on this work as well as two anonymous referees who considerably helped improve the paper. Excellent research assistance has been provided by Lawrie Harper-Simmonds. The research presented in this paper is based on a study undertaken for Thames Water Utilities Limited. The views expressed in this paper and any remaining errors are those of the authors alone.

[†]Corresponding author. Department of International Economics and Center for International Environmental Studies, Graduate Institute of International and Development Studies, Chemin Eugène Rigot 2, 1202 Geneva, Switzerland, E-mail: bruno.lanz@graduateinstitute.ch; and Chair for Integrative Risk Management and Economics, ETH Zurich, Switzerland.

[‡]Economics for the Environment Consultancy (eftec), London, UK.

1 Introduction

Well-functioning markets provide a way to aggregate information across firms and consumers. By observing consumer choices, firms make decisions about innovation and quality improvements to existing products, and associated investment priorities. For utility providers operating in regulated industries, such as energy or water and sewerage utilities, the role of consumer choice is typically diminished or non-existent. As a result, decisions about the appropriate level of service and associated price-setting regulation typically cannot be based on observed market behaviour.

In England and Wales, where the water industry is comprised of private regional monopolies, discrete choice experiments (DCE) (Louviere and Hensher, 1982; Louviere and Woodworth, 1983) are now widely applied to assess the value that consumers place on different aspects of services (e.g. the quality and reliability of water supply and sewerage services) and the management of natural resources (e.g. the ecological quality of rivers).¹ This follows from Willis et al. (2002), who suggested the use of cost-benefit analysis (CBA) to inform investment planning and price regulation.² The ‘exemplar’ study for valuing changes in the provision of water and sewerage services in this context is provided by Willis et al. (2005), who use the DCE approach to estimate marginal willingness to pay (WTP) for individual service attributes. This enables individual component values to be applied within CBA and inform investments across service attributes.

In this paper we examine the results of a DCE survey eliciting preferences for services of a water and sewerage company in England in line with the approach demonstrated in Willis

¹ Based on the assumption that consumers derive utility from characteristics of products (Lancaster, 1966; Rosen, 1974), a DCE simulates market transactions by constructing sets of alternative combinations of attributes (service characteristics) and requires respondents to select their most preferred alternative in a number of choice occasions. Trading-off various aspects of service provision the various services with changes in utility bills reveal preferences for independent changes in each attribute. Originally applied in the context of transportation (Ben-Akiva and Lerman, 1985), its application in recent years has encompassed marketing (Zwerina, 1997), health (Ryan, 1999), and the environment (Adamowicz et al., 1994).

² The Water Services Regulation Authority (Ofwat) regulates price-setting among water and sewerage companies in England and Wales. The amount by which consumer bills can change over time is determined by the five-yearly Periodic Review process, through which Ofwat scrutinises the proposed business plans of the regulated companies. It has been established that plans should be informed by the systematic comparison of the costs and benefits of service improvements (e.g. Ofwat, 2007), and explicitly take into account the preferences of customers (Ofwat, 2011).

et al. (2005). In common with many DCE applications, this requires survey respondents to repeatedly choose between a status quo (SQ) option describing a situation with no changes in the services considered and two alternatives. By definition, the price associated with the SQ is zero, so that it also represents a default alternative or outside good. The treatment of SQ choices, and in particular whether they reveal meaningful preferences from consumers, is an important issue for the regulator when scrutinising business plans that propose service improvements (and associated bill impacts) from already high levels.³ In the study we consider, the SQ was selected in 60 percent of all DCE choices, and 14 percent of respondents selected the SQ in all the choice occasions. As we review below, these patterns of responses are not uncommon in DCE studies.⁴

Our objective is to determine the extent to which SQ choices reflect the preferences of consumers, or alternatively contextual factors related to information provision and respondent cognition. Our survey instrument has a number of features geared towards identifying potential determinants of SQ choices. First, we elicit preferences for the SQ by recording satisfaction with the current level of service along with the general perception of how the service dimension impacts the household's daily activities. Second, we use a split sample design to vary the description of the SQ, and test whether information on the 'baseline' evolution of services and bills in the future affects the trade-off between the SQ option and the alternatives on offer. Third, we test whether the survey administration mode has an impact on SQ choices by administering the survey online and through face-to-face interviews where information provision can be monitored. Fourth, we define the attribute space to span both positive and negative domains, so that the 'price' of the hypothetical alternatives can be both positive (for service improvement) and negative (for a service deterioration). This implies that the SQ is not necessarily the least-cost

³ In the context of regulation in England and Wales, the widespread use of DCEs in the development of utilities' investment plans has generated significant scrutiny of stated preference methods by all stakeholders involved in the process (including the companies themselves, the regulator, and consumer representative groups, see UKWIR, 2010).

⁴ There exist a number of other empirical challenges associated with the use of DCEs. These include hypothetical bias (Diamond and Hausman, 1994; List, 2001), incentive compatibility of the choice format (Harrison, 2007), task complexity (Swait and Adamowicz, 2001), and preference 'anomalies' (Bateman et al., 2009; Day and Prades, 2010). Importantly, evidence from controlled field experiments have provided empirical support for the DCE approach to estimate marginal WTP in terms of hypothetical bias (List et al., 2006). Further, Vossler et al. (2012) have shown that DCEs can induce truthful revelation of preferences if choices are perceived to have a chance of affecting policy. While these issues are not directly studied in the present paper, the design of our survey instrument builds on these results to minimise their potential implications.

alternative. Finally, using a number of diagnostic questions, we control for respondents perceptions of the survey (e.g. the complexity of the choice tasks) and potential protest motivations.⁵

Importantly, evidence of a large number of SQ choices does not necessarily compromise the validity of DCEs as a preference revelation instrument. Indeed, as initially documented in Samuelson and Zeckhauser (1988), a tendency to stick with the SQ arises in many different settings and may thus be part of consumer's preferences.⁶ In turn, it represents important information for both companies and the regulator. However, SQ choices may also be used as a way to opt out of the exercise because of survey design effects, the cognitive burden imposed by the complexity of the choice task, or 'protest' against some aspect of the choice situation. Understanding whether SQ choices represent preferences or not is important because it underpins a utility-theoretic interpretation of stated choices. And given the growing use of DCEs to inform investment planning for the provision of water-related services and associated price regulation, the validity of SQ choices is directly relevant to the ex-post acceptability of company business plans.

While the design of the DCE survey is based on Willis et al. (2005), our analysis of SQ choice builds on a wider literature exploring SQ choices in DCEs as initially discussed in Adamowicz et al. (1998).⁷ Boxall et al. (2009) show that choice task complexity and the age of respondent are both positively related to the choice of the SQ. To some extent, Meyerhoff and Liebe (2009) confirm the role of perceived task complexity in determining SQ choices, but they also show that SQ choices reflect protest behaviour (see also Hanley et al., 2006, for a discussion). As in Meyerhoff and Liebe (2009), we use diagnostic questions to control for the perceived complexity and protest motives.

In their analyses of DCE SQ choices, both Boxall et al. (2009) and Meyerhoff and Liebe (2009) rely on interacting an alternative specific constant (ASC) with potential determinants

⁵ The 'protest' designation is borrowed from the contingent valuation literature (Freeman, 1986; Halstead et al., 1992) as a way to determine whether choices made by respondents reflect strategic motives triggered by some aspect of the elicitation method rather than preference for the good on offer.

⁶ A status quo 'bias' can be interpreted as a manifestation of loss aversion in a multiple good context (see Rabin, 1998, for a discussion). SQ choices may then be due to implications of losses appearing larger than the gains of other goods.

⁷ Two early papers by Hartman et al. (1990, 1991) studied the propensity of respondents to stay with the SQ in a contingent valuation survey evaluating the demand for electric service reliability. While these paper document what is called a SQ 'bias', they do not study the determinants of these choices.

of SQ choices, which limits the number of determinants that can be tested.⁸ In our analysis we directly model the probability of SQ choices, which allows us to assess a much wider range of factors. We employ this framework to provide direct evidence on the role of preference for the SQ, as measured by satisfaction with the experience service and importance of attributes in day-to-day activities.

The description of the SQ has been shown to impact SQ choices in the paper by Marsh et al. (2011). Allowing respondents to use their own description of the SQ, they find that those respondents who provided their own SQ were more likely to subsequently chose the SQ. This shows that respondents may be reluctant to opt away from outcomes that they know well, but also that perception of the SQ is heterogeneous. One explanation for this feature is that the SQ differs from experimentally defined alternatives in that it has been experienced (Scarpa et al., 2005, 2007; Hess and Rose, 2009). In our study, we compare a description of the SQ based on the current situation with a richer description providing detail about future evolution of services and bills, which potentially makes the SQ more comparable to the other alternatives.

Finally, while all the studies cited above focus on SQ choices in individual choice tasks, we also consider the determinants of 'serial' SQ choices, where a respondent selected the SQ in *all* choice occasions. The issue of serial SQ choices has been documented in Von Haefen et al. (2005), although to our knowledge there is currently no evidence about the determinants of serial SQ choices. We show that there exist differences between motivations underlying the individual and serial SQ choices, the latter being driven mainly by cognitive factors associated with information and complexity of the tasks. We conclude that individual SQ choices reveal relevant information about customer preferences. Therefore in settings where the SQ is a viable option from a policy or regulatory perspective there should be a way for customer to reveal such preferences. However, our results also suggest that there is a case for identifying serial SQ choices and determining whether they reflect cognitive or contextual factors. Sensitivity

⁸ The SQ ASC, together with an additional error term (or error component), controls for the role of unobserved sources of utility for the SQ and captures the fact that the perception of the SQ may systematically differ from the experimentally specified alternatives. However, while a positive SQ ASC to rationalise observed choices provides direct evidence about the prevalence of the SQ option in DCE choices (or its market share), it is inherently difficult to assess whether it reflects preferences for the SQ or whether it is a feature of the preference elicitation method.

analysis on the importance of these respondents for the wider results of the survey should be part of standard validity tests to increase confidence in the welfare estimates.

The remainder of this paper is structured as follows. Section 2 describes the experimental design from which our data are derived, provides details about the administration of the survey, and reports WTP-analysis of the DCE. Our empirical analysis of SQ choices is presented in Section 3. Discussion and conclusions are offered in Section 4.

2 Empirical framework: Valuing services of water utilities

Our data is sourced from a survey assessing customers preferences for water, sewerage and environmental services for a regulated company in England. This section describes the key features of our survey instrument and sample characteristics, and reports a WTP-space analysis of the DCE data.

2.1 Survey design

The questionnaire and DCE exercise follows the typical structure for a stated preference survey (Bateman et al., 2002) and was pre-tested through a series of focus groups, cognitive interviews and a pilot survey. As part of an introductory section, respondents are asked to rank priorities for improvements across a broad range of water, sewerage, environmental and consumer service areas. Respondents are then presented with the specific service areas covered by the DCE. Attributes presented in the DCE are reported in Table 1 (the full description that was made available to respondents is provided in Appendix A).

The attributes included in the DCE cover all the aspects of the company's services that would be affected by the investment schemes considered. These cover the quality of water supplied to households (interruption to supply; taste of tap water; water pressure; water use restrictions; water hardness), negative impacts of sewage treatment (sewage flooding; wastewater treatment odour) and environmental factors (river water quality; pollution incidents). In turn, given the inclusion of an attribute reflecting changes in bills, it is possible to estimate the value that a household places on changes to these services which can be compared to the costs of the investment schemes.

Table 1: Service attribute descriptions

Attribute name	Description	Units (Levels)
<i>Block 1</i>		
Interruption to supply	Number of properties affected each year by unexpected 3-6 hour loss of their water	100,000 properties per year (L-2: 2.5; L-1: 1.5; SQ: 1.2; L+1: 1.0; L+2: 0.5)
Taste of tap water	Number of complaints each year about taste or smell or colour of tap water	1,000 complaints per year (L-2: 7.2; L-1: 5.8; SQ: 4.9; L+1: 3.8; L+2: 2.7)
River water quality	River water quality assessments meeting 'good' or better	Percent per year (L-2: 66; L-1: 69; SQ: 72; L+1: 75; L+2: 78)
<i>Block 2</i>		
Water pressure	Number of properties affected each year by occasional low pressure	100,000 properties per year (L-2: 5.0; L-1: 2.5; SQ: 2.0; L+1: 1.5; L+2: 1.0)
Water use restrictions	Chance of restrictions on the use of water	Percent per year (L-2: 20.0; L-1: 10.0; SQ: 5.0; L+1: 3.3; L+2: 2.0)
Sewage flooding	Number of properties each year with a 1 in 10 year chance of internal sewage flooding	1,000 properties per year (L-2: 2.0; L-1: 1.4; SQ: 1.1; L+1: 0.8; L+2: 0.5)
<i>Block 3</i>		
Water hardness	Number of complaints received each year about the hardness of tap water	1,000 complaints per year (L-2: 1.8; L-1: 1.5; SQ: 1.2; L+1: 0.9; L+2: 0.6)
Pollution incidents	Number of pollution incidents each year	10 incidents per year (L-2: 5; L-1: 3; SQ: 2; L+1: 1; L+2: 0.5)
Wastewater treatment odour	Number of properties affected each year by smell from sewage treatment	1,000 properties per year (L-2: 10.0; L-1: 5.0; SQ: 2.0; L+1: 1.0; L+2: 0.5)
<i>All blocks</i>		
Bills	Change in annual household water and sewerage bill	UK£ per year (L-2: -30; L-1: -15; SQ: 0; L+1: +10; L+2: +20; L+3: +30)

One particular aspect of the service attributes is that they all have a public good component. Indeed customers on the same network cannot individually choose the level of service they would prefer, but rather experience an aggregate service level with the entire customer base. As highlighted by Willis et al. (2005) only a small fraction of customers experience a failure in any of the services (such as an interruption to supply). This implies that changes in probability of experiencing a service failure for a given household is typically small and difficult to comprehend (e.g. a 1 in 10,000 chance). As an alternative, respondents are asked to consider the overall change in the number of customers experiencing a particular service failure, so that services are valued across the entire base of customers. This then aligns with the level at which CBA is typically undertaken to establish the net present value of proposed investments in defined programme areas, such as reducing the number of properties affected by sewer flooding across the customer base.

For each service attribute, we specify two improved levels (L+1 and L+2) and two deterio-

rated levels (L-1 and L-2), while the bill attribute featured two reductions and three increases. Levels were determined in conjunction with the water company to reflect current state of services and improvements / deterioration that could realistically be considered.⁹ We elicit respondents' preferences for all attributes through the 'block' design used in the original study by Willis et al. (2005), where each block comprises three service attributes plus annual household bill. To mitigate the potential for an ordering effect the sequence of blocks is randomised across respondents.¹⁰ The total number of choice tasks is 12, with four tasks per block of attributes. In each choice task, respondents select their preferred option from the SQ and two alternative options where different levels of improvements / deterioration are specified. To determine the specific combinations of attribute levels in each choice task we apply a D-efficiency criteria (see e.g. Street and Burgess, 2007), which minimises the variance of the estimates and thus reduces the number of observation required to obtain a given level of precision in the estimation.

Turning to the specific issue of SQ choices, we directly elicit information about subjective perception of the SQ by asking two questions *prior* to the DCE exercise. First, for each attribute we elicit the satisfaction with the SQ level of service, as respondents unhappy with the current level would be most likely not to remain with the SQ (subject to cost). Second, we enquire whether each of the attributes considered have an impact on day-to-day household activities. We expect that, if changes in the attributes included in the choice set have little implications for a particular households, it is more likely to stay with the SQ.

The description of the SQ itself has been shown to influence incidence of SQ choices (Marsh et al., 2011). The usual description of the SQ tends to focus on just the experienced situation. However from the perspective of the respondent, it is important to understand investments that have already been committed to and that bills will increase over time to finance these in any case. Hence the description of SQ which is relevant for the company's decision-making purposes describes how service levels and bills will evolve over time due to already committed investments, so that the SQ option will differ from the currently experienced situation.

⁹ Figures are rounded for easier interpretation by respondents, something that was tested in focus groups and cognitive interviews.

¹⁰ In particular, this allows us to check that restricting the sample to the first block of services presented does not alter the welfare estimates.

To test how the description of the SQ influences SQ choices we use split-sample design. In the first subsample respondents are given a ‘basic’ description of the SQ which includes a general statement that water bills may increase to cover improvements to other aspects of services. We also include a conventional cheap talk script to mitigate potential hypothetical bias (Cummings and Taylor, 1999; List, 2001; Landry and List, 2007), remind respondents about income constraints, and state that their choices will be used to inform decision-making, which has been shown to be important for truthful preference revelation (Vossler et al., 2012). The ‘basic’ description of the SQ induces a difference between the SQ option and the two experimentally determined alternatives, as only the SQ has been experienced (Scarpa et al., 2005). The second subsample is used to contrast the basic SQ treatment with a ‘dynamic’ description of the SQ which describes future service and bill changes, and shifts the reference point from what is currently experienced to a situation which is yet to be experienced. Since this makes the three choice options more consistent, the dynamic SQ description is expected to reduce the prevalence of SQ choices.

A second split sample experiment is conducted to assess the effect of the survey administration mode. Online survey administration is becoming more prevalent due to its lower cost and increased internet access in the population (see Lindhjem and Navrud, 2011; Nielsen, 2011). However, the attention and care taken by respondents during the DCE exercise online cannot be directly monitored. In contrast, using a face-to-face computer aided personal interview (CAPI) approach means that an interviewer can directly monitor how the respondent behaves during the survey. Thus we administer the ‘dynamic’ SQ treatment both online and through a CAPI approach.

After completing the DCE exercise, follow-up questions elicit information on potential protest motives and respondents’ perceptions of the survey instrument, which were shown to be an important driver of SQ choices (Meyerhoff and Liebe, 2009). To identify protest motives, we include a diagnostic question focusing on three sources of protest. First, we consider the credibility of the choice scenario, eliciting how likely respondents thought that the services described would improve if bills were to increase. Second, we identify the main motivation for DCE choices. In particular the use of water bills as a payment vehicle can be problematic because

some respondents may perceive changes in bills as a way to raise the companies' revenues and profits rather than funding service improvements. Consequently we identify respondents whose main motivation for a SQ choice is that customers should not be charged for improvements, but rather that the company or government should pay for them. Finally, we identify respondents who felt the information provided in the survey and choice tasks is insufficient or unclear for them to make choices.

2.2 Survey administration and sampling procedure

The survey was administered in Autumn 2011 to approximately 1,500 household customers of the water and sewerage services company, split equally among the three sub-sample treatments (approx. 500 respondents each): 'basic' SQ (CAPI), 'dynamic' SQ (CAPI) and 'dynamic' SQ (online). The sub-sample size was determined based on the requirements of the experimental design for deriving statistically precise welfare estimates from the DCE and to ensure sufficient observations for comparison of results.¹¹

For the CAPI sub-samples, 48 sampling locations were selected reflecting the geographic extent of the water company region and relative split of urban and rural areas. Households were recruited within these sampling locations based on the random selection of a postcode starting point for interviewers, who then followed a set walk-pattern. The recruitment rate was around 1 completed survey for every 15 – 20 households approached. Each respondent was randomly allocated to either the basic or dynamic SQ treatment. Online respondents were recruited from an online panel. An invite to participate in the survey was sent to a sample of panel members whose home address was within the water company region. The response rate was around 1 completed survey for every 6 invites.

For each subsample, target quotas were specified in terms of respondent gender, age and household socio-economic status based on population statistics for the company customer base to ensure a representative sample. Both the quotas and the resulting sample composition are reported in Table 2. With some minor exceptions, the sub-sample characteristics are very close to

¹¹ Because we use an efficient experimental design, and have multiple observation per respondent, the sample of respondents is relatively modest. As we show below, the chosen sample size achieved the objective since all marginal WTP estimates are precisely estimated.

Table 2: Main socio-demographic characteristics and sampling quotas

Target	Customer base (sampling quotas)	CAPI basic (N=501)		CAPI dynamic (N=516)		Online dynamic (N=500)	
		Mean	Difference	Mean	Difference	Mean	Difference
<i>Gender</i>							
Female	0.52	0.53 (0.02)	-0.01 (0.59)	0.55 (0.02)	-0.03 (1.26)	0.52 (0.02)	0.00 (0.08)
<i>Socio-economic groups</i>							
A/B/C1	0.60	0.61 (0.02)	-0.01 (0.49)	0.61 (0.02)	-0.01 (0.37)	0.68 (0.02)	-0.08 *** (3.82)
<i>Age</i>							
18 - 29	0.24	0.20 (0.02)	0.04 (2.18)	0.24 (0.02)	0.00 (0.21)	0.24 (0.02)	0.00 (0.15)
30 - 44	0.33	0.33 (0.02)	0.00 (0.16)	0.32 (0.02)	0.01 (0.67)	0.33 (0.02)	0.00 (0.10)
45 - 59	0.21	0.27 (0.02)	-0.06 *** (2.71)	0.23 (0.02)	-0.02 (0.64)	0.21 (0.02)	0.00 (0.40)
60+	0.22	0.19 (0.02)	0.03 * (1.63)	0.22 (0.02)	0.00 (0.22)	0.22 (0.02)	0.00 (0.18)

Notes: Means are reported by sub-samples with standard deviations in parenthesis below. The column with “differences” reports differences between sub-samples means and quotas reported in the first column, with t-statistics reported in parenthesis below. ***, **, *: statistically significant at 1, 5 and 10 percent respectively.

the population statistics. Nevertheless, because some of these characteristics have been shown to be related to SQ choices (Meyerhoff and Liebe, 2009), we re-weight each subsample to be representative on observable characteristics.

2.3 Results: Discrete choice experiments

Our analysis of the DCE is standard and is also in line with Willis et al. (2005). As usual in the random utility framework, we start from the assumption that the choice of the preferred alternative results from utility maximisation. Formally, a respondent i chooses alternative j over j' in choice task t if the utility of j is greater than the utility of j' . The utility of alternative j is given by:

$$U_{ijt} = V_{ijt} + \epsilon_{ijt}, \quad i = 1, \dots, N, \quad j = 1, \dots, J, \quad t = 1, \dots, T,$$

where, from the viewpoint of the analyst, V_{ijt} and ϵ_{ijt} are the observed and unobserved parts of utility respectively. The function V_{ijt} is then a linear function of attributes' levels:

$$V_{ijt} = \sum_k \theta_{ik} X_{kjt} - \gamma_i P_{jt}$$

where the θ 's measure the marginal utility (or tastes) for improvements, X_{kjt} are levels of improvement specified in alternative j , γ is the marginal utility of money, and P_{jt} is the price of j .

To directly compare estimates across different blocks of services and split samples, we use a 'WTP-space' approach to estimation (Train and Weeks, 2005). Defining WTP for service improvements as $\beta_{ik} = \theta_{ik}/\gamma_i$, money-metric utility is obtained by rescaling V_{ijt} through γ_i . Furthermore, we control for the presence of individual-level heterogeneity in tastes with a random parameter specification (Revelt and Train, 1998): $\beta_{ik} \sim N(\bar{\beta}_k, \sigma_k)$ and $\gamma_i \sim \log N(\bar{\gamma}, \sigma_\gamma)$. We estimate the mean and the standard deviation parameters with simulated maximum likelihood. The utility of the SQ option also includes a dummy variable (the SQ ASC) and associated random parameter capturing the distribution of tastes for the SQ (aside from the utility derived from the level of service attributes defining the SQ).¹² A positive and statistically significant estimate for the SQ ASC is often interpreted as evidence of a SQ 'bias', as it indicates disproportionate choice of the SQ in relation to the (inferred) utility of the alternatives on offer.

Estimation results for each block and sub-samples are reported in Table 3.¹³ Explanatory variables are coded such that an increase implies an improvement in the level of service, and we thus expect a positive coefficient (WTP) on all service attributes. Results show that all estimates have the expected sign, and all are highly statistically significant except for the 'water use restrictions' attribute, which is statistically insignificant for two of the three subsamples. This finding is consistent with wider results from our survey that indicate the relatively minor

¹² This specification mirrors the error-component structure introduced by Scarpa et al. (2007) which allows the scale of the error variance to differ between the SQ option and the hypothetical alternatives. We favour the random parameter interpretation of the SQ ASC in this analysis as it provides direct evidence on preference heterogeneity for the SQ.

¹³ The model is estimated with simulated maximum likelihood, and we use 500 Halton draws to approximate the integral of the unconditional likelihood of each panel choices. Standard errors are clustered at the individual level to account for the fact that each respondent makes four different choices for each block of service.

Table 3: Discrete choice experiment: WTP-space estimation

Attribute	CAPI Basic SQ		CAPI Dynamic SQ		Online Dynamic SQ	
	Mean ($\bar{\beta}$)	Std-dev. (σ)	Mean ($\bar{\beta}$)	Std-dev. (σ)	Mean ($\bar{\beta}$)	Std-dev. (σ)
<i>Block 1</i>						
Interruption to supply	8.930*** (2.275)	7.971 (9.151)	10.30*** (2.581)	15.59*** (5.019)	9.696*** (1.183)	6.951*** (0.817)
Taste of tap water	6.245*** (1.055)	9.306*** (1.621)	7.899*** (1.829)	9.050*** (3.463)	6.785*** (0.681)	9.480*** (1.008)
River water quality	2.765*** (0.442)	4.123*** (0.664)	2.776*** (0.489)	3.457*** (0.824)	1.491*** (0.194)	2.165*** (0.296)
Bills	-3.200*** (0.129)	0.164 (0.210)	-2.995*** (0.693)	0.800 (0.762)	-1.364*** (0.467)	1.865*** (0.338)
SQ ASC	38.12*** (4.664)	43.01*** (5.972)	36.68*** (7.539)	40.29** (8.337)	20.18*** (1.822)	17.41*** (1.998)
N observations	2004		2064		2000	
N respondents	501		516		500	
Pseudo R ²	22.5		22.7		21.9	
<i>Block 2</i>						
Water pressure	9.442*** (1.762)	9.157*** (3.398)	7.762** (3.504)	5.645 (7.110)	5.368*** (0.645)	4.562*** (0.393)
Water use restrictions	-0.166 (0.190)	0.518 (1.205)	-0.183 (0.198)	0.0475 (0.134)	0.492*** (0.0922)	1.063*** (0.0841)
Sewage flooding	32.63*** (4.472)	22.26*** (6.144)	26.38*** (7.023)	18.79** (8.091)	20.28*** (1.647)	21.13*** (1.021)
Bills	-2.653*** (0.320)	1.103*** (0.327)	-2.032*** (0.275)	1.814*** (0.232)	-1.476*** (0.494)	1.824*** (0.432)
SQ ASC	34.15*** (4.593)	33.55*** (6.619)	25.40*** (3.955)	17.41*** (2.381)	12.99*** (1.909)	15.01*** (1.649)
N observations	2004		2064		2000	
N respondents	501		516		500	
Pseudo R ²	27.1		22.0		21.5	
<i>Block 3</i>						
Water hardness	8.859*** (2.111)	1.347 (2.707)	8.085*** (1.917)	9.670*** (1.260)	4.971*** (1.187)	10.48*** (1.758)
Pollution incidence	8.359*** (1.938)	14.49*** (3.632)	9.226*** (1.196)	10.45*** (1.775)	9.022*** (1.437)	9.097*** (1.622)
Wastewater treatment odours	5.550*** (1.493)	5.845*** (1.978)	3.720*** (0.406)	2.747*** (0.404)	2.136*** (0.290)	0.464* (0.255)
Bills	-3.581*** (0.272)	0.413 (0.415)	-2.005*** (0.740)	1.757*** (0.515)	-2.312*** (0.434)	1.159*** (0.402)
SQ ASC	54.95*** (10.78)	71.66*** (16.16)	28.16*** (5.253)	30.98*** (4.260)	13.54*** (2.231)	21.34*** (3.784)
N observations	2004		2064		2000	
N respondents	501		516		500	
Pseudo R ²	25.8		22.3		20.5	

Notes: All coefficients are marginal willingness to pay estimates in terms of the units specified for each attribute (Table 1). Standard errors are clustered at the individual level and reported in parenthesis. ***, **, *: statistically significant at 1, 5 and 10 percent respectively. Observations are weighted for representativeness of the customer base.

impact that temporary water use bans have on households.

Estimates of the SQ ASC are positive and highly statistically significant across all samples. This indicates that, all other things equal, respondents value the SQ over and above the value given to attributes defining the SQ alternative. In line with Scarpa et al. (2007) and Hess and Rose (2009), for all blocks and subsamples we observe that the standard-deviation estimates measuring individual-level deviations from the average is statistically significant. Given the magnitude of the standard deviation estimate, it implies that for some respondents the SQ option carries a negative utility weight, and these respondents would need to be compensated for staying with the SQ.

Estimates from different sub-samples reveal no systematic variations in either WTP estimates or taste heterogeneity, and any differences are small and statistically insignificant. However preferences for the SQ option vary. The SQ ASCs in the ‘basic’ SQ treatment are larger than those estimated from the ‘dynamic’ SQ treatment, and the SQ ASCs are lowest in the online subsample. As we document in next section, the size of the SQ ASC reflects the fact that the number of SQ choices (or market share of the SQ) is highest in the CAPI basic treatment and lowest in the online survey. Our subsequent analysis also examine whether differences in the prevalence of SQ choices are related to variations in the survey design (basic vs. dynamic SQ description) and administration (CAPI vs. online) or difference across respondents characteristics, in particular perception of the SQ.

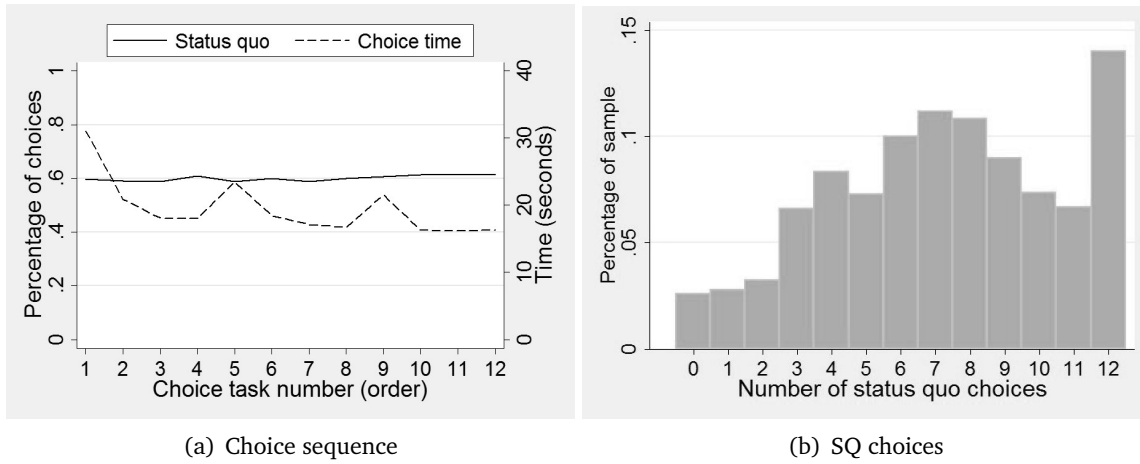
3 Analysis of status quo choices

This section examines the SQ choices by respondents and their motivations. We first provide descriptive statistics about SQ choices both within and across subsamples. We then turn to regression analysis to assess the determinants of SQ choices, first in individual choice tasks and then those for respondents who selected the SQ in all choice tasks, or serial SQ choices.

3.1 Descriptive statistics of status quo choices

Figure 1 presents the frequency of SQ choices across choice tasks. As panel (a) shows, around 60 percent of the combined sample choose the SQ in any given choice task. This is similar to the

Figure 1: Descriptive statistics on status quo choices (N=1,517)

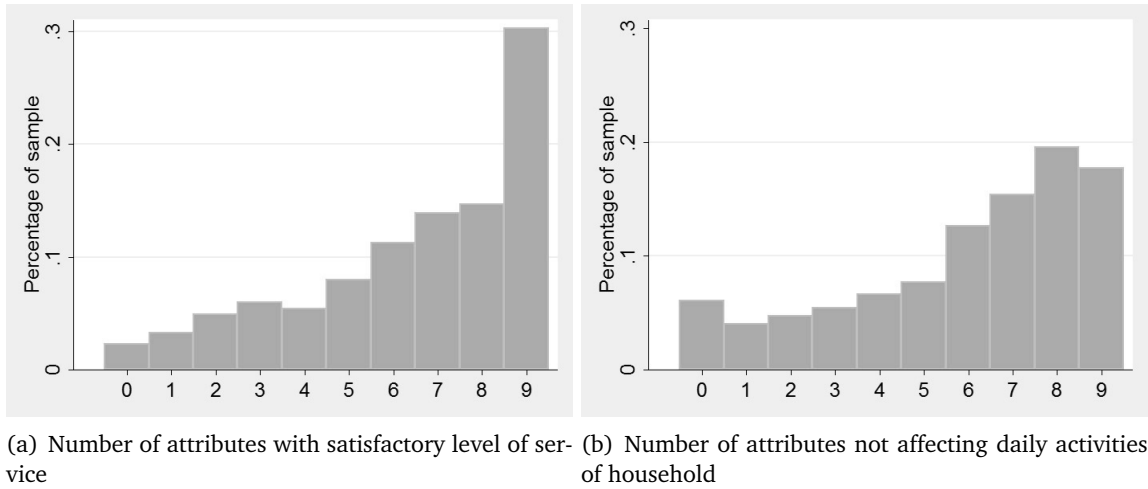


frequency of SQ choices reported in Hartman et al. (1990, 1991) and Boxall et al. (2009), but higher than the 25 percent reported in Hanley et al. (2006), for example. The proportion of SQ choices remains fairly constant across choice tasks and attribute blocks, although the average completion time for each task declines as the respondent progress through the exercise. As a new DCE block with different attributes is presented, the response time in each task significantly increases (see choices 5 and 9). This indicates a learning effect as a new block of attributes is introduced, which is consistent with Hensher et al. (2001), but this has little impact on the frequency of SQ choices.

In panel (b) of Figure 1, we report the number of SQ choices among the 12 tasks carried out by each respondent. About 14 percent of respondents opted for the SQ in all 12 choices, which is almost twice the frequency of respondents who selected the SQ alternative in all but one task. This suggests a discontinuity in the motivations underpinning serial SQ choices relative to idiosyncratic SQ choices. Note that this is significantly lower than Meyerhoff and Liebe (2009) who report between 50 and 60 percent of serial SQ choices, and Von Haefen et al. (2005) where 20 percent of respondents always chose SQ.

Figure 2 portrays the distribution of two key potential drivers of SQ choices, namely the satisfaction with the SQ level of services (panel a) and how much each service attribute affect the daily activities of the respondent's household (panel b). About 30 percent of respondents stated

Figure 2: Household perception of service provision (N=1,517)



that they were ‘happy’ with the current level of service for all attributes, while two percent of respondents indicated that all service attributes needed improvement. This potentially provides a very relevant explanation for the relatively high occurrence of SQ choices. Similarly, about 20 percent of respondents stated their daily activities were not affected ‘at all’ by any of the nine attributes. Note also that the Spearman’s rank correlation between the satisfaction with current service levels and the importance of attributes for daily activities is 56.4 percent.

Looking across sub-samples, Table 4 provides further descriptive evidence about SQ choices and perceptions of the SQ. As already hinted by the WTP-space analysis of the DCE data, the prevalence of SQ choices is highest in the online sample and lowest in the CAPI basic subsample. The difference with the CAPI dynamic subsample is statistically significant at the 10% and 5% level respectively. In the online sample we also observe a lower prevalence of serial status quo choices, and the difference with the CAPI dynamic subsample is statistically significant.

Turning to the perception of the SQ, we observe that dissatisfaction with the SQ level of services and the importance of these services in daily activities is, on average, lower in the online subsample. The differences in average satisfaction between CAPI dynamic and online samples are highly statistically significant for *all* attributes. Therefore, the larger prevalence of SQ choices in the online subsample may be due to different perception of the SQ rather than due to the administration mode itself. We examine this in detail in the next section. More

Table 4: Status quo choices and preferences for the SQ across subsamples

Variable	CAPI dynamic (N=516)		CAPI basic (N=501)		Online dynamic (N=500)	
	Mean		Mean	Difference	Mean	Difference
Status quo choice (1=yes)	0.60 (0.01)		0.63 (0.01)	-0.03* (-1.88)	0.56 (0.01)	0.04** (2.35)
Serial status quo choices (1=yes)	0.15 (0.02)		0.16 (0.02)	-0.01 (-0.53)	0.10 (0.01)	0.05** (2.28)
<i>Happy with the current provision</i>						
Interruption to supply (1=yes)	0.83 (0.02)		0.85 (0.02)	-0.02 (-0.67)	0.75 (0.02)	0.08** (3.18)
Taste of tap water (1=yes)	0.82 (0.02)		0.82 (0.02)	0.00 (0.12)	0.72 (0.02)	0.10*** (3.65)
River water quality (1=yes)	0.68 (0.02)		0.65 (0.02)	0.02 (0.77)	0.45 (0.02)	0.23*** (7.38)
Water pressure (1=yes)	0.84 (0.02)		0.87 (0.01)	-0.03 (-1.41)	0.76 (0.02)	0.08*** (3.19)
Use restrictions (1=yes)	0.85 (0.02)		0.86 (0.02)	0.00 (-0.12)	0.74 (0.02)	0.11*** (4.41)
Flooding events (1=yes)	0.84 (0.02)		0.85 (0.02)	-0.01 (-0.34)	0.70 (0.02)	0.14*** (5.21)
Water hardness (1=yes)	0.61 (0.02)		0.58 (0.02)	0.03 (1.10)	0.30 (0.02)	0.31*** (10.33)
Pollution events (1=yes)	0.71 (0.02)		0.69 (0.02)	0.02 (0.54)	0.42 (0.02)	0.29*** (9.56)
Wastewater treatment odour (1=yes)	0.78 (0.02)		0.80 (0.02)	-0.01 (-0.56)	0.63 (0.02)	0.15*** (5.23)
Number of attributes (0-9)	6.96 (0.11)		6.96 (0.11)	0.00 (0.02)	5.47 (0.11)	1.49*** (9.53)
<i>Not affected by changes</i>						
Interruption to supply (1=yes)	0.85 (0.02)		0.84 (0.02)	0.01 (0.52)	0.55 (0.02)	0.29*** (10.58)
Taste of tap water (1=yes)	0.69 (0.02)		0.71 (0.02)	-0.02 (-0.71)	0.46 (0.02)	0.23*** (7.60)
River water quality (1=yes)	0.71 (0.02)		0.73 (0.02)	-0.02 (-0.63)	0.48 (0.02)	0.23*** (7.49)
Water pressure (1=yes)	0.73 (0.02)		0.70 (0.02)	0.03 (1.13)	0.43 (0.02)	0.30*** (10.05)
Use restrictions (1=yes)	0.85 (0.02)		0.85 (0.02)	0.00 (0.07)	0.48 (0.02)	0.38*** (13.61)
Flooding events (1=yes)	0.87 (0.01)		0.86 (0.02)	0.01 (0.56)	0.73 (0.02)	0.15*** (5.77)
Water hardness (1=yes)	0.38 (0.02)		0.35 (0.02)	0.03 (0.93)	0.11 (0.01)	0.27*** (10.54)
Pollution events (1=yes)	0.83 (0.02)		0.85 (0.02)	-0.02 (-0.79)	0.56 (0.02)	0.27*** (9.47)
Wastewater treatment odour (1=yes)	0.80 (0.02)		0.80 (0.02)	0.00 (0.07)	0.61 (0.02)	0.19*** (6.67)
Number of attributes (0-9)	6.71 (0.11)		6.68 (0.10)	0.03 (0.21)	4.40 (0.13)	2.31*** (14.12)

Notes: Means are reported by sub-samples with standard deviations in parenthesis below. The column with “differences” reports differences in means between the respective sub-samples and the CAPI dynamic subsample reported in the first column, with t-statistics reported in parenthesis below. ***, **, *: statistically significant at 1, 5 and 10 percent respectively. Observations are weighted for representativeness of the customer base.

importantly, this reveals some imbalance in subsamples even though the samples are balanced in terms of observable socio-demographic characteristics. As already documented in Nielsen (2011), this is potentially evidence that online survey administration may be associated with a sample selection issue. However, imbalances in the preferences for the SQ across subsamples have no systematic impacts on WTP for the different services, as we have shown in Table 3 above. Therefore, while these differences in preferences for the SQ may explain the differences in the frequency of SQ choices, these do not have a systematic impact on the demand for service improvements. This finding is consistent with Nielsen (2011) who finds a potential for non-response bias in online survey but no statistically significant differences in mean and median WTP values.

3.2 Econometric analysis of SQ choices

This section reports the econometric analysis of SQ choices. First, we analyse the drivers of SQ choices in each choice occasion and across all DCE blocks. Second, we model the probability that respondents systematically choose the SQ option in all 12 choice tasks (serial SQ choices), reflecting systematic non-participation in the DCE exercise. In both cases we treat the choice between SQ and non-SQ options as a separate process from the choice between the two other alternatives on offer. As compared to a multinomial logit estimation reported in Boxall et al. (2009) and Meyerhoff and Liebe (2009) where the SQ ASC is interacted with relevant variables, focusing on the probability of SQ choices allows us to pool the data from different DCE blocks. Furthermore, this modelling approach significantly increases the number of degrees of freedom and remains consistent with the random utility framework.

3.2.1 Assessing the determinants of status quo choices

To analyse the probability of observing a SQ choice in each choice occasion, we pool data on the 12 choice tasks. Thus for each respondent i we observe 12 binary SQ outcomes (indexed by t) across the three blocks of attributes (indexed by l), and we exploit variations both within and

across respondents using a binary logit model:

$$Prob(SQ_{it} = 1 | X_{ikjt}, P_{ijt}, Z_{ilt}, W_{is}) = f \left(\alpha + \sum_{ikj} \delta_{kj} X_{ikjt} + \sum_j \lambda_j P_{ijt} + \sum_l \omega_l Z_{ilt} + \sum_s \rho_s W_{is} \right)$$

where $f(\cdot)$ is the logit function, X_{kjt} and P_{jt} represent the attributes' levels of each alternative, Z_{ilt} is a set of indicator variables to identify each DCE block, W_{is} is a vector of individual and survey characteristics that remain constant across choice tasks but vary across individuals, and α , δ , λ , ω and ρ are parameters to be estimated from the data. Note that the variables included in W_{is} are mainly informed by the work of Boxall et al. (2009) and Meyerhoff and Liebe (2009), and control for the perception of the survey including its complexity, protest motives, and key socio-demographic variables. In addition, we control for the satisfaction with the SQ level of services and the importance of these services in daily activities,¹⁴ and add a number of variables quantifying the impact of elicitation, namely one indicator variable for each subsample and a variable capturing the order of the choice task in the sequence (Day and Prades, 2010).

Results from the estimation are reported in Table 5. Standard errors are clustered at the individual level to account for the fact that each respondent makes twelve choices in total. Marginal improvements in service attributes display the expected sign. An increase in service level makes that particular option more attractive, reducing the probability of selecting the SQ (and the opposite for the bills). The only exception is 'water use restrictions', for which we obtain an unintuitive result for one of the two alternatives. We also observe that if one of the two non-SQ alternatives included a reduction in bills, the probability of choosing the SQ declines by about 5 percent. This confirms evidence of gains-loss asymmetry in DCEs reported elsewhere (e.g. Lanz et al., 2010; Viscusi and Huber, 2012), and suggests that a significant share of SQ choices is driven by respondents who select the cheapest option.

We find that both measures of preferences for the SQ generally have a positive association with the probability of SQ choice, and these are quantitatively important determinants of SQ choices. The marginal effect for the taste of tap water attribute is 12.5 percent, suggesting that

¹⁴ Interaction terms between these two variables were also tested but did not yield further insights.

Table 5: Binary logit model for individual status quo choices (marginal effects)

DCE block	DCE attribute	Marginal improvement		Happy with the current provision (=1)	Not affected by changes (=1)
		Alternative 1	Alternative 2		
Block 1 (baseline)	Interruption to supply	-0.02 (0.01)	-0.02* (0.01)	0.05* (0.02)	-0.004 (0.02)
	Taste of tap water	-0.02*** (0.005)	-0.02*** (0.005)	0.03 (0.02)	0.12*** (0.02)
	River water quality	-0.01*** (0.002)	-0.01*** (0.002)	0.04** (0.02)	-0.02 (0.02)
Block 2 (=1) -0.01 (0.03)	Water pressure	-0.03*** (0.01)	0.002 (0.01)	0.03 (0.03)	0.04 (0.02)
	Water use restrictions	0.002* (0.001)	-0.01*** (0.001)	0.05* (0.03)	0.02 (0.02)
	Sewage flooding	-0.13*** (0.02)	-0.10*** (0.02)	0.02 (0.02)	0.03 (0.03)
Block 3 (=1) -0.02 (0.03)	Water hardness	-0.01 (0.01)	-0.01 (0.01)	0.02 (0.02)	0.05** (0.02)
	Pollution incidents	-0.02*** (0.01)	-0.03*** (0.005)	0.03* (0.02)	0.04 (0.03)
	Wastewater treatment odour	-0.01*** (0.002)	-0.01*** (0.002)	0.04 (0.02)	0.02 (0.03)
All blocks	Bills	0.004*** (0.0003)	0.003*** (0.0003)	- -	- -
	Negative bill amount (=1)	-0.05** (0.02)	- -	- -	- -
<i>Elitication effects</i>	Basic SQ (=1)	0.03* (0.02)	<i>Survey perception</i>	Interesting (=1)	-0.02 (0.02)
	Online survey (=1)	0.002 (0.02)		Complicated (=1)	0.003 (0.03)
	Task order effect (1-12)	0.002** (0.001)		Educational (=1)	-0.03 (0.02)
<i>Respondent characteristics</i>	Age (years)	0.001*** (0.0005)	<i>Protest motives</i>	Changes unlikely (=1)	0.03* (0.02)
	Gender (female=1)	0.05*** (0.01)		Customers shouldn't pay (=1)	0.03* (0.02)
	Univ. education (=1)	-0.02 (0.02)		Insufficient information (=1)	0.06* (0.04)
	Socio-econ groups (A/B/C1=1)	0.005 (0.02)			

Notes: N observations = 18,204; N respondents = 1,517. Pseudo-R²: 4.3%. Marginal effects evaluated at the sample mean. Standard errors are clustered at the individual level and reported in parenthesis. ***, **, *: statistically significant at 1, 5 and 10 percent respectively. Observations are weighted for representativeness of the customer base.

this attribute is a key determinant of SQ choices.

Variations in the survey instrument reveal that the basic SQ description is associated with a higher probability of SQ choices, all other things equal. This is as expected, since the basic description of the SQ relates more directly to the situation experienced by the respondents whereas the dynamic description makes the three alternatives more comparable, thus facilitating trade-offs across alternatives. The indicator variable for the online subsample is very near zero and not statistically significant. This is despite the fact that descriptive statistics (and WTP-

space analysis) suggested that online respondents were more likely to select the SQ option. A key finding here is that the indicator variable for the survey mode becomes irrelevant only when we control for preferences for the SQ, namely whether respondents are happy with the currently experienced service level and whether attributes does not affect daily activities of the households. In other words, the prevalence of SQ choices among online respondents is related to their preferences for the SQ rather than to the survey mode.¹⁵

We find some evidence that the probability of SQ choices increases with the a variable measuring the choice task order from 1 to 12, which implies that the occurrence of SQ choices increases as the exercise progresses. However, the difference in SQ choice probability between the first and the last choice, all other things equal, is about 1 percent. We also find no evidence that the probability of SQ choices is affected by the block of services per se, as shown by the coefficients associated with Block 2 and 3.

We further classify individual characteristics into three different categories. First, subjective perception of the survey is not found to be associated with the probability of SQ choices, whereas we would have expected the probability of SQ choice to increase for respondents who found the exercise ‘complicated’ (Boxall et al., 2009). Second, coefficients for variables identifying protest behaviour are all quantitatively large and statistically significant ($p < 0.1$), implying that protest motives increase the probability of SQ choices. The finding that the perceived complexity of the exercise is irrelevant when controlling for protest motives is in line with Meyerhoff and Liebe (2009). Third, among socio-demographic characteristics, age and gender are both correlated with different probability of SQ choices, all other things equal, while education and socio-economic classification are not.

Based on these results, we can simulate the probability of SQ under different instances. For example, if both hypothetical alternatives offered maximum improvement and zero cost, and that all respondents were dissatisfied with the experience service level and that all attributes mattered for respondents’ day-to-day activities, the probability of SQ choices would decline

¹⁵ Because of imbalances across subsamples in terms of preferences for the SQ, it may be the case that the impact of preferences for the SQ is driven by the inclusion of the online subsample. As we report in Appendix B, the main conclusions from our analysis, and in particular the importance of preferences for the SQ, are preserved if we restrict the sample to just the CAPI subsamples.

from an observed 60 percent to 6.8 percent.¹⁶ If we further remove the effect of protest motives by setting the three associated dummies to zero, SQ choices would only represent 6.1 percent of choices. Thus our model accounts for the greater proportion of SQ choices, demonstrating that only a very small fraction of SQ choices are not explained by preferences for services on offer.

3.2.2 Assessing the determinants of serial SQ choices

We now analyse the probability of serial SQ choices, which is directly associated with respondents' participation in the DCE exercise. In this setting, the dependent variable is an indicator variable equal to one if the respondent choose the SQ option in all tasks, zero otherwise. As above, we use a binary logit model:

$$Prob\left(\sum_t SQ_t = 12 | W_{is}\right) = f\left(\alpha + \sum_s \rho_s W_{is}\right)$$

where we only exploit individual and survey level characteristics W , and α , ρ are parameters to be estimated from the data. Results of the estimation are reported in Table 6.

First, we find that the perception of the SQ, as measured by the number of attributes whose provision is satisfactory and the number of attributes perceived not to affect the respondent's household daily activities, has a positive impact on serial SQ choices. These effects are of practical and statistical significance, supporting the view that for some respondents serial SQ choices reflect preferences for the SQ.¹⁷ In contrast, we do not find any difference in the probability of serial SQ choices among the different subsamples. Thus differences in serial SQ choices observed for CAPI and online subsamples are explained by differences in preferences for the SQ rather than by the survey administration mode.¹⁸

Turning to the perception of the survey, we find that self-assessed complexity of the choice exercise correlates with serial SQ choices, so that the more complex the exercise is perceived to be, the greater the probability of serial SQ choices. Thus in our analysis the effect of complexity

¹⁶ All other variables included are at the sample mean.

¹⁷ Note that we do not find evidence supporting the presence of non-linearities for these variables.

¹⁸ As for the individual SQ choices restricting the sample to CAPI respondents only does not affect our main conclusions (see Appendix C).

Table 6: Binary logit model for serial SQ choices (marginal effects)

<i>DCE attributes</i>	Happy with the current provision (number of attributes: 0-9)	0.01** (0.004)			
	Not affected by changes (number of attributes: 0-9)	0.01** (0.004)			
<i>Survey characteristics</i>	Basic SQ (=1)	0.01 (0.02)	<i>Survey perception</i>	Interesting (=1)	0.003 (0.02)
	Online survey (=1)	-0.01 (0.02)		Complicated (=1)	0.06** (0.03)
				Educational (=1)	-0.05** (0.02)
<i>Respondent characteristics</i>	Age (years)	0.001** (0.001)	<i>Protest motives</i>	Changes unlikely (=1)	0.03 (0.02)
	Gender (female=1)	0.02 (0.02)		Customers shouldn't pay (=1)	0.003 (0.02)
	Higher education (=1)	-0.02 (0.02)		Insufficient information (=1)	0.10* (0.05)
	Socio-econ groups (A/B/C1=1)	0.02 (0.02)			

Notes: N = 1,517; Pseudo-R²: 5.3%; Marginal effects evaluated at the sample mean. Robust standard errors are reported in parenthesis. ***, **, *: statistically significant at 1, 5 and 10 percent respectively. Observations are weighted for representativeness of the customer base.

only appears through serial SQ choices. Conversely, respondents who stated that the survey was educational are less likely to choose the SQ option in all tasks. Among respondents' characteristics, only age is positively correlated with serial SQ choices, a factor which was already found to be related to individual SQ choices. Protest motives are associated with a higher probability of serial SQ choice, although concerns about insufficient information is the only protest motive which is statistically significant at conventional levels.

As for individual SQ choices, we can simulate the probability of serial SQ choices in different settings. From around 14 percent in the sample, if respondents were unhappy with all attributes and affected by changes in all attributes on offer, the probability of serial SQ choices would drop to 4.2 percent.¹⁹ If we further set indicator variables for protest motives to zero, the serial SQ choice probability declines to 3.7 percent. Thus the share of serial SQ choices not accounted by our model is fairly modest, although the role of preferences appears somewhat smaller compared to individual SQ choices.

¹⁹ Again all other variables are kept at the sample mean.

4 Discussion and conclusion

The aim of this paper has been to examine the determinants of SQ choices in DCEs in the context of regulation of the water industry in England and Wales. Given the recent emphasis on the use of DCE to support the companies business plans and the associated price regulation, understanding whether the large share of SQ choices (around 60 percent in our sample) reflects preferences for attributes of the SQ or derives from the elicitation method is important. Indeed from a regulatory perspective SQ choices that do not reflect preferences may be excluded from the sample, as per the treatment of protest motives in contingent valuation (e.g. Halstead et al., 1992).

Using the approach demonstrated by Willis et al. (2005) to quantify customer preferences for water and sewerage services, we elicited satisfaction with the SQ and importance of attributes in day-to-day activities to measure preferences for the SQ, and varied both the description of the SQ and the survey administration mode. Our results show that the perception of the SQ plays a major role in explaining SQ choices. Importantly, this perception and the relevance of attributes are also key determinants of serial SQ choices. Taken together, these results imply that for many respondents the SQ is valuable because they are satisfied with current service levels or because they do not feel not directly affected by changes in the provision of service attributes. SQ choices thus hold important information for price regulation, and are not unexpected given that service levels in England and Wales are already high. From this perspective, our results provide further evidence to support the use of DCE as a preference elicitation instrument (see also List et al., 2006; Vossler et al., 2012) and, in the present case, as a tool to account for customer preferences in regulated sectors, particularly in relation to investment planning and price setting. More generally, in DCE applications where the SQ is a viable choice from a policy or regulatory perspective, the inclusion of an alternative representing the SQ is important to avoid constraining customers in selecting their preferred alternative.

While our results show that SQ choices do not necessarily reflect poor study design, or lack of respondent understanding or care, we also confirm that a share of SQ choices are not related to preferences. In fact this issue is critical for serial SQ choices, which are found to be mainly driven by respondents finding the provision of information insufficient or that the

choice tasks were too complex. This suggests that responses from customers who systematically selected the SQ option *and* signalled issues with the content of the survey should be treated with caution. Indeed from a regulatory perspective choices of these individuals do not represent truthful revelation of preferences and may bias welfare estimates that are relevant for policy decisions. Our results call for a careful treatment of these respondents, and identification serial SQ choices that reflect cognitive or contextual factors should be part of standard validity tests to increase confidence in the results.

With regards to survey design, we find that a ‘dynamic’ description of the SQ reduces the occurrence of SQ choices, as it likely makes SQ and non-SQ options more comparable. Indeed in many DCEs the SQ is the only experienced alternative and respondents may select the SQ to avoid uncertainties associated with experimentally designed alternatives. Another source of SQ choices comes from the fact that the SQ is typically the cheapest option, whereas allowing for reduction in bills reduces the probability of SQ choices, all other things equal. These two effects can easily be accounted for at the stage of the survey design. From a regulatory perspective, providing respondents with a description of the SQ that reflects the future evolution of services and bills, as well as an opportunity to select reductions in services and bills, thus appears important.

Finally, we provided new insights on the difference between online and CAPI survey administration. In particular, we found that online respondents were more likely to select the SQ in the DCE exercise, which translated in a higher SQ ASC. We also found very clear evidence that respondents to the online survey were, on average, less satisfied with the SQ. While these differences did not have a systematic impact on WTP estimates, it explains the finding that the SQ ASC is higher in the analysis of DCE choices. For our purposes, we found that once we controlled for preferences for the SQ, there was no statistically significant differences in the probability of SQ choices across sampling modes. Therefore survey administration mode in itself does not appear to affect SQ choices. Nevertheless imbalances in the composition of the sample indicates that there might be a sample selection effect affecting unobservables, as was already hinted by Nielsen (2011), and thus calls for further research on this topic.

Appendix A Attribute descriptions presented to respondents

Unexpected interruptions to supply	<p>Interruptions to the water supply can occur without any warning or notice due to pipes bursting or equipment failing. This could happen at any time of the day or night.</p> <p>Affected properties have no water supply for a temporary period (between 3 to 6 hours).</p>
Taste, smell or colour of tap water	<p>Some customers experience recurring incidents of unpleasant taste or smell or discolouration of their tap water.</p> <p>Properties can be affected over a period of time (e.g. a week) and running the tap for several minutes does not remove the taste, smell or discolouration.</p>
River water quality	<p>All rivers in Europe should meet environmental quality standards that aim to improve the status of natural habitats.</p> <p>The quality of rivers and stream habitats for wildlife and use by people can be affected by treated wastewater discharged from sewage treatment works.</p>
Low water pressure	<p>Low water supply pressure can be due to a property's location or a reduction in mains pressure. It affects the use of taps, showers and boilers - for example it can take up to 5 times longer to fill a bowl of water.</p> <p>'Occasional' means the water pressure is not low all the time but becomes lower in several one-off incidents over a year.</p>
Water use restrictions	<p>During long periods of dry weather and low rainfall - usually summer months - it may be necessary to restrict the use of water by customers.</p> <p>Restrictions on uses like watering gardens and parks, washing cars and windows, and filling swimming pools conserves water for essential uses (drinking, cooking and washing) and helps maintain river levels to protect water habitats and wildlife.</p>
Sewage flooding	<p>Very occasionally, problems with sewers particularly during periods of heavy rainfall can result in untreated sewage escaping from sewers.</p> <p>Internal sewage flooding means that the sewage water gets into houses and other buildings. This could be in basement areas or in ground level rooms.</p>
Hardness of tap water	<p>Hardness of tap water refers to the mineral content of water. This is not a concern for public health and the water is safe to drink, but it does affect the use of kettles, boilers and other appliances through 'scaling'.</p> <p>This is an issue all year round for customers in the affected areas.</p>
Pollution incidents	<p>The Environment Agency for England and Wales sets standards for discharges of sewage, in order to protect the wildlife in rivers and the use of rivers by people.</p> <p>Occasionally the sewerage system is affected by mechanical failures, blockages and heavy rain. This can result in the discharge of untreated sewage to rivers. This may affect the wildlife in rivers for a temporary period of time.</p>
Wastewater treatment odour	<p>Sewers and sewage treatment works can be the source of complaints about unpleasant smell and nuisance, such as flies, even when they are operating normally.</p> <p>Properties close to treatment works are most affected and problems can also be worse in warmer weather.</p>

Appendix B Individual SQ choices in the CAPI subsamples

Table B1 reports the results of Section 3.2.1 restricted to the CAPI subsamples: CAPI basic and CAPI dynamic. We report marginal effects for a logit model evaluated at the mean of the sample. Standard errors are clustered at the individual level to account for the fact that each respondent makes twelve choices in total.

Table B1: Binary logit model for individual status quo choices (marginal effects)

DCE block	DCE attribute	Marginal improvement		Happy with the current provision (=1)	Not affected by changes (=1)
		Alternative 1	Alternative 2		
Block 1 (baseline)	Interruption to supply	-0.02 (0.01)	-0.01 (0.01)	0.02 (0.03)	-0.02 (0.03)
	Taste of tap water	-0.02*** (0.01)	-0.02*** (0.01)	0.04 (0.03)	0.13*** (0.02)
	River water quality	-0.01*** (0.00)	-0.01*** (0.00)	0.11*** (0.02)	-0.02 (0.03)
Block 2 (=1) -0.02 (0.05)	Water pressure	-0.03*** (0.01)	0.01 (0.01)	0.02 (0.04)	0.07** (0.03)
	Water use restrictions	0.002* (0.001)	-0.005*** (0.001)	0.02 (0.04)	0.05 (0.04)
	Sewage flooding	-0.14*** (0.02)	-0.13*** (0.02)	0.08** (0.04)	-0.02 (0.04)
Block 3 (=1) -0.03 (0.04)	Water hardness	-0.02* (0.01)	-0.01 (0.01)	0.01 (0.03)	0.04 (0.03)
	Pollution incidents	-0.01 (0.01)	-0.02*** (0.01)	0.07*** (0.03)	0.04 (0.04)
	Wastewater treatment odour	-0.01*** (0.00)	-0.01*** (0.00)	0.04 (0.03)	0.03 (0.04)
All blocks	Bills	0.004*** (0.000)	0.002*** (0.000)	- -	- -
	Negative bill amount (=1)	-0.04 (0.03)	- -	- -	- -
<i>Elitication effects</i>	Basic SQ (=1)	0.03* (0.02)	<i>Survey perception</i>	Interesting (=1)	0.02 (0.02)
	Online survey (=1)	-		Complicated (=1)	0.01 (0.03)
	Task order effect (1-12)	0.002* (0.001)		Educational (=1)	-0.02 (0.03)
<i>Respondent characteristics</i>	Age (years)	0.001** (0.001)	<i>Protest motives</i>	Changes unlikely (=1)	0.04** (0.02)
	Gender (female=1)	0.04** (0.02)		Customers shouldn't pay (=1)	0.02 (0.02)
	Univ. education (=1)	-0.02 (0.02)		Insufficient information (=1)	0.08* (0.04)
	Socio-econ groups (A/B/C1=1)	-0.003 (0.020)			

Notes: N observations = 12,204; N respondents = 1,017. Pseudo-R²: 4.8%. Marginal effects evaluated at the sample mean. Standard errors are clustered at the individual level and reported in parenthesis. ***, **, *: statistically significant at 1, 5 and 10 percent respectively. Observations are weighted for representativeness of the customer base.

Appendix C Serial SQ choice for the CAPI subsamples

Table C1 reports the results of Section 3.2.2 restricted to the CAPI subsamples: CAPI basic and CAPI dynamic. We report marginal effects for a logit model evaluated at the mean of the sample.

Table C1: Binary logit model for serial SQ choices (marginal effects)

<i>DCE attributes</i>	Happy with the current provision (number of attributes: 0-9)	0.011* (0.006)			
	Not affected by changes (number of attributes: 0-9)	0.010 (0.006)			
<i>Survey characteristics</i>	Basic SQ (=1)	0.01 (0.02)	<i>Survey perception</i>	Interesting (=1)	0.05* (0.03)
	Online survey (=1)	-		Complicated (=1)	0.08* (0.04)
				Educational (=1)	-0.05 (0.03)
<i>Respondent characteristics</i>	Age (years)	0.001 (0.001)	<i>Protest motives</i>	Changes unlikely (=1)	0.04 (0.03)
	Gender (female=1)	0.02 (0.02)		Customers shouldn't pay (=1)	0.01 (0.02)
	Higher education (=1)	-0.03 (0.03)		Insufficient information (=1)	0.14* (0.08)
	Socio-econ groups (A/B/C1=1)	-0.002 (0.025)			

Notes: N = 1,017; Pseudo-R²: 4.7%; Marginal effects evaluated at the sample mean. Robust standard errors are reported in parenthesis. ***, **, *: statistically significant at 1, 5 and 10 percent respectively. Observations are weighted for representativeness of the customer base.

References

- Adamowicz, W., J. Louviere, and M. Williams (1994) "Combining stated and revealed preference methods for valuing environmental attributes," *Journal of Environmental Economics and Management*, 26, pp. 271 – 292.
- Adamowicz, W., P. Boxall, M. Williams, and J. Louviere (1998) "Stated preference approaches for measuring passive use values: choice experiments and contingent valuation," *American Journal of Agricultural Economics*, 80, pp. 64 – 75.
- Bateman, I., R. Carson, B. Day, W. Hanemann, N. Hanley, T. Hett, M. Jones-Lee, G. Loomes, S. Mourato, E. Ozdemiroglu, D. Pearce, R. Sugden, and J. Swanson (2002) *Economic Valuation with Stated Preference Techniques: A Manual*: Edward Elgar Publishing, Cheltenham.
- Bateman, I., B. Day, A. Jones, and S. Jude (2009) "Reducing gain – loss asymmetry: A virtual reality choice experiment valuing land use change," *Journal of Environmental Economics and Management*, 58 (1), pp. 106 – 118.
- Ben-Akiva, M. and S. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*: The MIT Press, Cambridge/Massachusetts.
- Boxall, P., W. Adamowicz, and A. Moon (2009) "Complexity in choice experiments: Choice of the status quo alternative and implications for welfare measurement," *The Australian Journal of Agricultural and Resource Economics*, 53, pp. 503 – 519.
- Cummings, R. and L. Taylor (1999) "Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method," *American Economic Review*, 89 (3), pp. 649 – 665.
- Day, B. and J. P. Prades (2010) "Ordering anomalies in choice experiments," *Journal of Environmental Economics and Management*, 59, pp. 271 – 285.
- Diamond, P. and J. Hausman (1994) "Contingent valuation: is some number better than no number," *Journal of Economic Perspectives*, 8 (4), pp. 45 – 64.
- Freeman, A. (1986) "On assessing the state of the arts of the contingent valuation method valuing environmental changes," in R. Cummings, D. Brookshire, and W. Schulze eds. *Valuing environmental goods: An assessment of the contingent valuation method*: Rowman and Allanheld Publishers, Totwa NJ.
- Halstead, J., A. Luloff, and H. Stevens (1992) "Protest bidders in contingent valuation," *North-eastern Journal of Agricultural and Resource Economics*, 21 (2), pp. 160 – 169.
- Hanley, N., R. E. Wright, and B. Alvarez-Farizo (2006) "Estimating the economic value of improvements in river ecology using choice experiments: An application to the water framework directive," *Journal of Environmental Management*, 78 (2), pp. 183 – 193.
- Harrison, G. (2007) "Making choice studies incentive compatible," in B. Kanninen ed. *Valuing Environmental Amenities Using Stated Choice Studies. A Common Sense Approach to Theory and Practice*, pp. 67 – 110: Springer, The Netherlands.
- Hartman, R., M. Donae, and C. Woo (1990) "Statusquo bias in the measurement of value of service," *Resources and Energy*, 12 (2), pp. 197 – 214.

- (1991) “Consumer rationality and the status quo,” *Quarterly Journal of Economics*, 106 (1), pp. 141 – 162.
- Hensher, D., P. Stopher, and J. Louviere (2001) “An exploratory analysis of the effects of numbers of choice sets in designed choice experiments: An airline choice application,” *Journal of Air Transport Management*, 7, pp. 373 – 379.
- Hess, S. and J. Rose (2009) “Should reference alternatives in pivot design SC surveys be treated differently?” *Environmental and Resource Economics*, 42 (3), pp. 297–317.
- Lancaster, K. J. (1966) “A new approach to consumer theory,” *Journal of Political Economy*, 74, pp. 132 – 157.
- Landry, C. and J. List (2007) “Using ex ante approaches to obtain credible signals for value in contingent markets: Evidence from the field,” *American Journal of Agricultural Economics*, 89 (2), pp. 420 – 429.
- Lanz, B., A. Provins, I. Bateman, R. Scarpa, K. Willis, and E. Ozdemiroglu (2010) “Investigating willingness to pay - willingness to accept asymmetry in choice experiments,” in S. Hess and A. Daly eds. *Choice Modelling: the State-of-the-Art and the State-of-Practice*: Emerald Group Publishing.
- Lindhjem, H. and S. Navrud (2011) “Are internet surveys an alternative to face-to face interviews in contingent valuation?” *Ecological Economics*, 70 (9), pp. 1628 – 1637.
- List, J. (2001) “Do explicit warnings eliminate the hypothetical bias in elicitation procedures? evidence from field auctions for sportscards,” *American Economic Review*, 91 (5), pp. 1498 – 1507.
- List, J. A., P. Sinha, and M. H. Taylo (2006) “Using choice experiments to value non-market goods and services: Evidence from field experiments,” *The B.E. Journal of Economic Analysis & Policy – Advances*, 6, p. Article 2.
- Louviere, J. and D. Hensher (1982) “On the design and analysis of simulated choice or allocation experiments in travel choice modelling,” *Transportation Research Record*, 890, pp. 11 – 17.
- Louviere, J. and G. Woodworth (1983) “Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data,” *Journal of Marketing Research*, 20, pp. 350 – 367.
- Marsh, D., L. Mkwara, and R. Scarpa (2011) “Do respondents’ perceptions of the status quo matter in non-market valuation with choice experiments? an application to new zeland fresh-water streams,” *Sustainability*, 3, pp. 1593 – 1615.
- Meyerhoff, J. and U. Liebe (2009) “Status quo effect in choice experiments: Empirical evidence on attitudes and choice task complexity,” *Land Economics*, 85 (3), pp. 515 – 528.
- Nielsen, J. S. (2011) “Use of the internet for willingness-to-pay surveys. a comparison of face-to-face and web-based interviews,” *Resource and Energy Economics*, 33 (1), pp. 119 – 129.
- Ofwat (2007) “Further ofwat guidance on the use of cost-benefit analysis for PR09.” Letter to all regulatory directors of water and sewerage companies and water only customers, Fiona Pethick, Ofwat Director of Corporate Affairs. http://www.ofwat.gov.uk/pricereview/pr09phase1/pr09phase1letters/ltr_pr0908_cbaguide [Accessed November 2014].

- (2011) “Involving customers in price setting - ofwat’s customer engagement policy statement.” http://www.ofwat.gov.uk/future/monopolies/fpl/customer/pap_pos20110811custengage.pdf [Accessed November 2014].
- Rabin, M. (1998) “Psychology and economics,” *Journal of Economic Literature*, 36 (1), pp. 11 – 46.
- Revelt, D. and K. Train (1998) “Mixed logit with repeated choices: Households’ choices of appliance efficiency level,” *Review of Economics and Statistics*, 80, pp. 647–657.
- Rosen, S. (1974) “Hedonic prices and implicit markets: Product differentiation in pure competition,” *Journal of Political Economy*, 82, pp. 34 – 55.
- Ryan, M. (1999) “A role for conjoint analysis in technology assessment in health care?” *International Journal of Technology Assessment in Health Care*, 15, pp. 443 – 457.
- Samuelson, W. and R. Zeckhauser (1988) “Status quo bias in decision making,” *Journal of Risk and Uncertainty*, 1, pp. 7 – 59.
- Scarpa, R., S. Ferrini, and K. Willis (2005) “Performance of error component models for status-quo effects in choice experiments,” in R. Scarpa and A. Alberini eds. *Applications of Simulation Methods in Environmental and Resource Economics*: Springer, Dordrecht.
- Scarpa, R., K. Willis, and M. Acutt (2007) “Valuing externalities from water supply: Status quo, choice complexity and individual random effects in panel kernel logit analysis of choice experiments,” *Journal of Environmental Planning and Management*, 50 (4), pp. 449 – 466.
- Street, D. and L. Burgess (2007) *The Construction of Optimal Stated Choice Experiments: Theory and Methods*: Wiley-Interscience.
- Swait, J. and W. Adamowicz (2001) “The influence of task complexity on consumer choice: a latent class model of decision strategy switching,” *Journal of Consumer Research*, 28, pp. 135 – 148.
- Train, K. and M. Weeks (2005) “Discrete choice models in preference space and willingness-to-pay space,” in R. Scarpa and A. Alberini eds. *Applications of simulation methods in environmental and resource economics*: Dordrecht: Springer, pp. 1–16.
- UKWIR (2010) “Review of cost-benefit analysis and benefit valuation.” Report Ref. No. 10/RG/07/18.
- Viscusi, W. K. and J. Huber (2012) “Reference-dependent valuations of risk: Why willingness-to-accept exceeds willingness-to-pay,” *Journal of Risk and Uncertainty*, 44 (1), pp. 19 – 44.
- Von Haefen, R., M. D.M., and W. Adamowicz (2005) “Serial nonparticipation in repeated discrete choice models,” *American Journal of Agricultural Economics*, 87 (4), pp. 1061 – 1076.
- Vossler, C. A., M. Doyon, and D. Rondeau (2012) “Truth in consequentiality: Theory and field evidence on discrete choice experiments,” *American Economic Journal: Microeconomics*, 4(4), pp. 145 – 171.
- Willis, K., P. McMahon, G. D. Garrod, and N. Powe (2002) “Water companies’ service performance and environmental trade-offs,” *Journal of Environmental Planning and Management*, 45 (3), pp. 363 – 379.

Willis, K., R. Scarpa, and M. Acutt (2005) "Assessing water company customer preferences and willingness to pay for service improvements: A stated choice analysis," *Water Resource Research*, 41, p. W02019.

Zwerina, K. (1997) *Discrete choice experiments in marketing: Use of priors in efficient choice designs and their application to individual preference measurement*: Physica-Verlag, Heidelberg.