THE GRADUATE INSTITUTE GENEVA

CENTRE FOR INTERNATIONAL ENVIRONMENTAL STUDIES

Research Paper 36 | **2015**

AVERTIVE EXPENDITURES, ENDOGENOUS QUALITY PERCEPTION, AND THE DEMAND FOR PUBLIC GOODS: AN INSTRUMENTAL VARIABLE APPROACH

B. LANZ

Avertive expenditures, endogenous quality perception, and the demand for public goods: An instrumental variable approach*

Bruno Lanz†

This version: October 2015

Abstract

In response to the perceived quality of a public good, households may choose to incur avertive expenditures as a substitute to its aggregate provision, thereby revealing an (inverse) demand function. When unobserved heterogeneity affects both perceived quality and avertive behavior, identification of the demand function is plagued by a problem of endogeneity. In this paper, I propose to use a first stage model of perceived quality as a function of objective quality to recover unbiased and microconsistant estimates of marginal willingness to pay for the provision of the public good. The approach can be applied when people have well-formed perceptions of the quality of the good, a prerequisite for the avertive expenditures method. I illustrate the approach with data on avertive expenditures for two qualitative aspects of household tap water networks: water hardness and aesthetic quality in terms of taste and odor.

Keywords: Public good provision; Avertive expenditures; Revealed preferences; Perceived quality; Objective quality; Water hardness; Aesthetic water quality; Water regulation.

JEL Codes: H4, L9, Q2, Q5, D1.

*I thank Allan Provins for his insight and collaboration on this project, as well as Ian Bateman, Diane Dupont, Scott Reid, participants at EPFL and 2014 WCERE meeting for their comments and suggestions. Excellent research assistance was provided by Martin Baker and Lawrie Harper-Simmonds. I also thank the participating water companies for allowing me to access their data. The views expressed in this paper and any remaining errors are mine.

†Department of Economics and Centre for International Environmental Studies, Graduate Institute of International and Development Studies, Geneva, Switzerland; Chair for Integrative Risk Management and Economics, ETH Zurich; Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology. Maison de la Paix P1-633, Chemin de Eugène Rigot 2, 1202 Geneva, Switzerland; Tel: +41 22 908 62 26; email: bruno.lanz@graduateinstitute.ch.

1 Introduction

There is an established practice of using household's avertive behavior and associated expenditures to estimate the private benefits of public good provision, and more generally price non-market goods and externalities (Courant and Porter, 1981; Harford, 1984; Harrington and Portney, 1987). In this framework households face an exogenous supply of a public good, but they can select their preferred level of provision by incurring costly actions. By observing how avertive expenditures vary with the objective level of provision, it is possible to identify an inverse demand schedule for the public good, also known as the marginal willingness to pay (WTP) schedule or valuation function (Cameron and James, 1987; Cameron, 1988). This has led to a number of empirical applications mainly focusing on the value of morbidity and mortality risks reductions using variations in air pollution (e.g Gerking and Stanley, 1986; Deschenes et al., 2012) and water pollution (e.g Smith and Desvousges, 1986; Harrington et al., 1989; Abdalla et al., 1992; Larson and Gnedenko, 1999; McConnell and Rosado, 2000; Abrahams et al., 2000; Yoo and Yang, 2000; Zivin et al., 2011).

A fundamental requirement to apply the avertive expenditure approach is that households observe the objective quality of provision. However, it is the *perceived* failure to reach the privately desired provision level that will determine avertive behavior and expenditures (Dickie and Gerking, 1996; Abrahams et al., 2000). In instances where individuals have heterogeneous perception of the public good provision, failure to control for the relationship between objective provision and perceptions may generate biased estimates, and potentially misleading policy recommendations (Whitehead, 2006). Heterogeneity affecting differences between perceived and objective quality is particularly pervasive when evaluating changes in risks (Slovic, 2000), as the same objective risk level may induce very different behavioral response depending on preferences and households' situation. To account for such differences, a number of studies have employed a measures of perceived provision to identify marginal WTP estimates (Um et al., 2002; Rosado et al., 2006; Jakus et al., 2009; Schram et al., 2010; Dupont and Jahan, 2012).

While using perceived quality in the valuation function potentially generates better estimates of marginal WTP, it also raises two potential issues. First, perceived quality combines information about objective provision and preferences, so that marginal WTP estimates no longer represent an inverse demand function. In fact, from a welfare theoretic perspective, interpretation of these figures is unclear and raises questions in their use to inform the socially optimal level of provision. Second, the perception of quality (as measured through survey questions) is itself an outcome, being a function of household characteristics and experiences (Danielson et al., 1995; Dupont, 2005; Nauges and van den Berg, 2009). Therefore perceived quality is potentially endogenous in an econometric sense. In particular, when unobserved factors affect both avertive behavior and quality perception, identification of marginal WTP with variations in perceived quality is likely biased because of endogeneity (Whitehead, 2006; Nauges and van den Berg, 2009; Orgill et al., 2013; Adamowicz et al., 2014; Bontemps and Nauges, 2015).

In an attempt to address these two issues, this paper proposes to combine information on perceived (subjective) and objective provision level. In particular, I model the relationship between objective quality and subjective perception in a first stage regression, and then include instrumented subjective quality as part of the valuation function. The exclusion restriction relies on an assumption that objective quality is exogenous and affects the demand for substitute products only through perceived quality. Thus on the one hand the valuation function accounts for the fact that the driver of choices (perceived quality) is only indirectly determined by objective quality, and addresses potential endogeneity of perceived quality. On the other hand, the first stage regression quantifies the relationship between subjective and objective provision, and can be used to obtain theoretically valid marginal WTP estimates that relate to the policy-relevant objective level of provision.[1]

To illustrate this approach, I employ data from a survey administered in England and Wales eliciting expenditures on substitutes for two characteristics of tap water supply, namely water hardness and the aesthetic quality in terms of taste and odor of tap water (see Lanz and

[1] In some settings, measures of objective and subjective quality may only be weakly related (see Orgill et al., 2013, for example). It is thus important to document the empirical validity of the relationship between objective and subjective quality measures as part of the estimation. However, note that the avertive behavior approach relies on agents knowing the objective provision level so that a failure of the first stage relationship may either indicate that the empirical measure of objective quality is not relevant to the decision-maker or that applying the avertive behavior method is not appropriate altogether.

Provins, 2014, for a comprehensive description of the survey and its wider results).[2] Aside from recording avertive expenditures by households in relation to hardness and aesthetic quality of tap water, the survey provides information about perceptions of tap water quality. In this context, one potential source of unobserved heterogeneity that could give rise to the endogeneity of perceived quality is preference learning. There is ample evidence that experience with public goods affects its valuation (e.g. Whitehead et al., 1995; Cameron and Englin, 1997; Czajkowski et al., 2014), and in the case of avertive behavior substitute products will provide consumers with an alternative experience of the good. This sort of experiences will then likely affect both perceived quality and expenditures on substitute products. The issue is particularly relevant for water hardness, since the benefits of softening devices can only be observed if a household actually uses such products.[3]

In order to obtain an objective measure of quality, survey respondents are matched to highly disaggregated regional data at the level of the water supply zones (WSZ).[4] Specifically, the instruments considered are WSZ-level data on average water hardness and on the rate of customer complaints to water service suppliers concerning aesthetic quality. Overall, results from the analysis confirm that perceived quality is endogenous, which implies that marginal WTP relating to perceived quality is biased towards zero; instrumenting perceived quality with objective quality yields marginal WTP estimates that are around around two times higher for water hardness and three time higher for aesthetic quality. However, while a good measure of objective quality is available for water hardness, and our analysis supports the use of the proposed estimation strategy, we emphasize that results for perceived aesthetic quality are subject to a number of caveat. Indeed our objective measure of aesthetic quality is in fact derived from subjective measures (complaints), and it is possible to imagine cases in which

[2] Hard water, via scaling, can damage and significantly reduce the lifetime of water-using appliances implying notable costs to households. While the level of hardness is a characteristic of the raw water source and is mainly determined by the geology of the area from which it is abstracted (specifically the presence of calcium and magnesium in aquifers), it can be mitigated by investments by the water utility in treatment plants or at the individual household level.

[3] For aesthetic characteristics most consumer would have had access to bottled water providing close-to-perfect aesthetic quality, and from a preference learning perspective the case for an endogeneity problem is weaker. Nevertheless filtering devises also involve a learning component.

[4] WSZ are geographically small areas in which water is provided from the same source (or sources) and usually a single water treatment works. The quality of water can then be assumed to be similar for all the customers in the same WSZ.

they are endogenous (e.g. if dissatisfaction spreads by word of mouth). Therefore, results on aesthetic quality should be treated with caution.

This paper is closely related to two recent studies tackling the issue of endogenous quality perception in the context of an avertive behavior model.[5] First, Adamowicz et al. (2014) study water consumption choices as a function of perceived mortality risk using a household survey in Canada. As part of the survey, they elicit perceived mortality risk from skin cancer and use it as an instrument for perceived mortality risk from water consumption. They find that a failure to instrument perceived quality implies a marginal value of risk reduction around 70 percent lower as compared to the values estimated with an instrumental variable procedure. Second, Bontemps and Nauges (2015) study the role of perceived health impacts (satisfied or not satisfied) on the (binary) decision of households to drink water directly from tap in Australia, Canada, and France. They instrument the perception variable with data from a preceding survey (with a different sample), aggregated at the regional level, on the proportion of household who drank water from the tap and concerns concern about general water pollution in that region.[6] They provide statistical evidence that the perception variable is endogenous, leading to a downward bias, although the magnitude of the bias in their preferred specification is only around 10 percent. Relative to the instrumental variable strategies used in these papers, the one proposed here is theoretically motivated, as the first stage can be used to identify marginal WTP in relation to the objective level of provision, and hence can be applied to inform the socially optimal provision level.

The remainder of this paper is structured as follows. Section 2 provides the analytical framework supporting the estimation. Section 3 describes the empirical illustration. Section 4 concludes.

---

[5] See Whitehead (2006) and Orgill et al. (2013) for a treatment of endogeneity in contingent valuation studies. Neither of these studies consider objective quality as an instrument.

[6] Note that the same authors have shown previously that household are more likely to drink water from the tap if they perceive the quality of their local environment to be high (see Bontemps and Nauges, 2009).

2 Theory and econometric identification

This section provides the framework supporting the econometric estimation of an avertive expenditure function. I start with a standard avertive expenditure model using the household production function framework of Becker (1965), Grossman (1972) and Courant and Porter (1981). The present exposition draws from Bartik (1988) who links avertive expenditures to a 'bid function' for quality improvements, and lays the ground for the empirical analysis using a valuation or WTP function introduced by Cameron and James (1987) and Cameron (1988). I then show how the use of self-reported perceived quality may result in biased marginal WTP estimates. The framework naturally suggests an estimation strategy using objective quality as an instrument for subjective quality.

2.1 Avertive expenditures model

Consumers have preferences represented by utility function $U(X, q)$ where X is a numeraire good (with price normalized to one) and q the perceived quality of a public good *consumed*. The variable q is also called 'personal' quality, as it potentially differs from the perceived quality of aggregate provision of the public good, denoted by w. In the case of tap water supply, q represents perceived quality of water that consumers drink, while w measures how consumers rate the quality of water supplied at the tap by their water utility. The key difference between these two variables is that consumers treat w as exogenous, in the sense that it is not a choice variable, whereas private goods may be purchased to improve personal quality q, and therefore drive a difference between q and w. Moreover, while *objective* quality of provision $\overline{w}$ is by definition very similar for all consumers (e.g. water from the same treatment plant), w will be heterogeneous for example because of household characteristics and past experiences. We come back to the relationship between w and $\overline{w}$ below.

The cost of improved perceived quality $q \geq w$ is given by an avertive cost function $AVC(q, w)$, with $\partial AVC/\partial q > 0$ and $\partial AVC/\partial w < 0$. Given income M and budget constraint $M = X + AVC(q, w)$, consumers will purchase substitute products until the desired quality level q^* is reached. Formally, maximizing the function $U(X, q)$ subject the budget constraint above yields the following first order conditions for an interior solution:

$$\partial U/\partial X = \lambda\,; \quad \partial U/\partial q = \lambda\, \partial AVC/\partial q \tag{1}$$

where λ is a Lagrange multiplier associated with the budget constraint. Hence $\frac{\partial U/\partial q}{\partial U/\partial X} = \partial AVC/\partial q$, which is the standard condition equating the marginal rate of substitution to the ratio of prices.

As suggested by Courant and Porter (1981), expenditures on substitute products provide information about the value that consumers put on changes in the provision of the public good $\overline{w}$ (measuring objective aggregate provision). This can be shown formally by deriving a compensating surplus measure defined as the difference in expenditures, keeping utility constant, that is equivalent to the change in the quality of the public good. In the present setting, changes in aggregate provision affect behavior through a function $w = f(\overline{w}; S)$, where S is a vector of individual characteristics. The function $f(\cdot)$ captures how individuals form perceptions about aggregate provision of the public good (see Dickie and Gerking, 1996, for example). For a marginal change in $\overline{w}$ an expression for the compensating surplus can be obtained by setting the total derivative of the indirect utility function $v = V[P, w, M]$ to zero, where P is a vector of market prices. At the optimum (X^*, q^*), the indirect utility function is equal to the Lagrangian of the problem, and the envelope theorem implies:

$$\begin{aligned} V(w, M) &= U(X^*, q^*) + \lambda(M - X^* - AVC(q^*, w)) \quad \Leftrightarrow \\ -\frac{dM}{d\overline{w}}\bigg|_{v^0} &= \frac{\partial V/\partial \overline{w}}{\partial V/\partial M} = \frac{\partial AVC(q, w)}{\partial w} \cdot \frac{\partial f(\overline{w}; S)}{\partial \overline{w}} \end{aligned} \tag{2}$$

where v^0 refers to a reference (constant) utility level, and prices in the vector P are assumed constant and hence omitted for simplicity.

For non marginal changes or if there are discontinuities in the avertive cost function, q^* will generally also change as $\overline{w}$ (and w) increase. Under the conditions discussed in Bartik (1988), variations in the avertive cost function will generally provide a lower bound for the compensating surplus associated with a change in $\overline{w}$.

2.2 Econometric identification of marginal willingness to pay

Following Bartik (1988) I now define a bid function $B(\cdot)$ for a marginal quality change $\Delta\overline{w} = \overline{w}^1 - \overline{w}^0$ as the difference in expenditures $e(\cdot)$ evaluated at different provision levels:

$$B(\Delta\overline{w}; Z) = e(f(\overline{w}^0; S), v^0, Z) - e(f(\overline{w}^1; S), v^0; Z) \tag{3}$$

where Z represents a vector of consumer characteristics (including for example household income) which potentially differs from S. The bid function is also known as a valuation or WTP function, and at the optimum it coincides with the definition of the compensating surplus underlying the avertive cost function (Cameron and James, 1987; Cameron, 1988). Hence for a consumer i we have that observed avertive expenditures *AVC** is equal to the bid function associated with perceived quality, indirectly affected by $\overline{w}$:

$$AVC_i^* = B(f(\overline{w}_i; S_i); Z_i) + \varepsilon_i \tag{4}$$

where ε_i is a residual term capturing unobservable components of the bid function. The objective is then to specify an empirical counterpart for the valuation function and identify the marginal effect of $\overline{w}_i$ on avertive expenditures AVC_i, providing an estimate of the welfare impact associated with changes in the provision of the public good.

As discussed by Whitehead (2006), a failure to control for variations in perceived quality in the bid function may generate biased marginal WTP estimates. WTP is a function of perceived quality of the public good, which is likely to vary across individuals even when objective quality is constant. Moreover, changes in objective quality are mediated by the function $f(\cdot)$, and can be expected to have a heterogeneous impact on perceived quality. While the use of perceived quality to identify marginal WTP can address this issue, it gives rise to an endogeneity problem if some unobserved factors influence both perceived quality and WTP. This can be seen by introducing an error term in the perception function:

$$w_i = f(\overline{w}_i, S_i) + \eta_i \tag{5}$$

It follows that when $\rho = corr(\varepsilon_i, \eta_i) \neq 0$, estimated partial effect of w_i on AVC_i will be biased.

Given the framework developed above, a natural instrument for perceived quality is a measure of objective quality. Indeed equation (5) readily provides an exclusion restriction, since objective quality affects WTP only through perceived quality. Avertive expenditures and quality perception are modeled as a two-stage process, whereby changes in an objective measure of provision first translates into an improved perceived quality, and improved perceived quality reduces avertive expenditures.

While quite general in principle, it is important to note that the exclusion restriction might be violated in some settings. Perhaps the most obvious challenge to its generality occurs when the provision of the public good of interest enters the location decision of households. For example, Bayer et al. (2009) and Tra (2010) provide evidence that household sort across neighborhood based on prevailing air quality. Thus for such particularly salient public goods, the objective provision level may itself be endogenous, which further complicates identification of marginal WTP values through an avertive behavior approach. However, in the context which we next consider, namely the choice of drinking water quality, this is unlikely to be an issue.[7]

2.3 Estimation strategy

The aim of the econometric analysis is to estimate a valuation function (4) and identify the effect of $\overline{w}_i$ on AVC_i through w_i. I treat avertive expenditures as a corner solution outcome, assuming that households with zero expenditures optimally chose this amount, and I use a tobit model to represent the conditional expectation of expenditures. Observed avertive expenditures are $AVC_i = \max(0, AVC_i^*)$, where AVC_i^* is a latent variable. AVC_i^* is then modeled as a function of quality rating w_i and a vector of household characteristics Z_i:

$$AVC_i^* = \gamma w_i + \beta' Z_i + \varepsilon_i \tag{6}$$

where γ represents the marginal WTP for changes in perceived quality, β is a vector of parameters and $\varepsilon \sim N(0, \sigma^2)$ is an error term.

[7] We note that for water quality choices, the exclusion restriction might also be violated in instances where household can substitute between different water sources (such piped supplies, wells, surface water), as objective quality could be related to source characteristics. In applications focusing on tap water systems where source choice does not enter, this is less of an issue.

Determination and potential endogeneity of perceived quality w_i is accounted for by means of a simultaneous-equation tobit model (Smith and Blundell, 1986; see also Nelson and Olson, 1978, and Amemiya, 1979). Specifically, the valuation function and the equation determining perceived quality are jointly estimated as:

$$\begin{cases} AVC_i^* = \gamma w_i + \beta' Z_i + \varepsilon_i \\ w_i = \phi \overline{w}_i + \alpha' S_i + \eta_i \\ AVC_i = \max(0, AVC_i^*)\,, \quad \rho = corr(\varepsilon_i, \eta_i) \end{cases} \tag{7}$$

where ϕ and α are parameters to be estimated from the data, and ρ measures the correlation between error terms and is also estimated from the data.

Expected avertive expenditures conditional on the vector of covariates (w_i, Z_i), denoted by $E(AVC_i \,|\, AVC_i \geq 0, w_i, Z_i)$, can be decomposed into two parts:

$$E(AVC_i \,|\, AVC_i \geq 0, w_i, Z_i) \;=\; P(AVC_i > 0 \,|\, w_i, Z_i) \cdot E(AVC_i \,|\, AVC_i > 0, w_i, Z_i) \tag{8}$$

Marginal WTP estimates for improved quality perception, $\frac{\partial E(AVC_i \,|\, AVC_i \geq 0, w_i, Z_i)}{\partial w_i}$, comprise both changes in the decision of whether or not to incur expenditures on substitutes (or a change in the fraction of households with positive avertive expenditures), as measured by $P(AVC_i > 0 \,|\, w_i, Z_i)$, and changes in the average amount spent by households who decide to incur expenditure, denoted $E(AVC_i \,|\, AVC_i > 0, w_i, Z_i)$. Marginal WTP estimates, and more generally marginal effects, are a highly non-linear function of the set of estimated parameter (γ, β) and are evaluated at a given value of the vector of covariates. Moreover, the policy-relevant marginal WTP with respect to objective quality is obtained through the chain rule:

$$\frac{\partial E(AVC_i \,|\, AVC_i \geq 0, w_i, Z_i)}{\partial \overline{w}_i} \;=\; \frac{\partial E(AVC_i \,|\, AVC_i \geq 0, w_i, Z_i)}{\partial w_i} \cdot \frac{\partial w_i}{\partial \overline{w}_i} \tag{9}$$

where $\frac{\partial w_i}{\partial \overline{w}_i} = \phi$ is obtained from the first stage regression. Therefore, this specification provides an indirect micro-consistent approach to evaluate the marginal WTP for a change in the objective provision level.

3 Empirical application

In this section, I illustrate the use of the estimation strategy in the context of tap water quality choices. First I describe data collection and provide summary statistics. I then turn to the econometric estimation of the valuation function across alternative specifications. Finally results for total avertive expenditures and marginal WTP estimates for perceived and objective quality changes are provided.

3.1 Survey instrument and data sources

Empirical results reported in this paper are mainly based on a survey of customers of water companies in England and Wales. The survey instrument was developed through several stages, including a national omnibus survey to determine the set of avertive behavior carried out by households and an online pilot with a sample of approximately 200 respondents.

The final survey is structured as follows. Following a screening question on the respondent's responsibility for paying household bills, a set of warm-up questions focus on the composition of the respondent's household (number of people and age groups) and their consumption of tap water for drinking and other uses. Information on the consumption of substitutes for tap water quality is then elicited, including water filters (e.g. a jug/kettle, tap/under sink filter, fridge dispenser), bottled water, squash and cordial, water softener devices (that remove calcium and magnesium ions from tap water) and other products (e.g. tablets, powders and coils).

Given their use of substitute products, respondents are then filtered to follow-up questions in which they indicate the specific product types they use/purchase, their substitute uses of these products (e.g. drinking, food preparation, washing, watering plants, etc.) and their expenditures, including one-off amounts, regular amounts, and the frequency of purchases. Following this, respondents are asked to indicate why their household uses substitutes to tap water, including reasons related to the aesthetic quality of tap water, water hardness, health concerns, advice from water utility (e.g. do not drink notices), medical professionals and other sources (media, advertising, etc.), and other motivations (e.g. convenience, temperature of tap water). This allows determining which expenditures are related to avertive behavior.

The survey then asks respondents about specific experiences and perceptions related to the use of substitutes, including experience of problems with aesthetic quality of tap water (e.g. chlorine taste, musty taste, cloudy appearance), the quality of tap water (e.g. hardness, impurities) and health issues (e.g. risk of illness, contaminants and pollutants, lead in supply pipes). Respondents provide a rating of the tap water supply at their home on a 1 to 5 scale, in terms of its taste, odor, appearance, hardness, and overall quality. The survey concludes with questions about the respondent's household including how long they have lived at their current address, their annual water bill amount, their own health status and the health status of others in their household.

Survey responses are augmented with data measuring objective quality of tap water. The data is taken from mandatory reporting requirement by water companies to the Drinking Water Inspectorate for England and Wales. Objective quality measures relates to individual WSZ, which are geographically small areas with an average population of around 30,000 supplied from the same source of treated water.[8] In each WSZ, an objective measure of water hardness is given by the average mg of calcium carbonate ($CaCO3$) per liter. For aesthetic quality, I use data on the number of customer complaints relating to the taste and odor of tap water together with total population in each WSZ to compute an annual rate of complaints related to taste and odor of tap water. Respondents are then matched to their WSZ through their home postcode.

3.2 Data and summary statistics

The survey was administered online in November and December 2012 via a panel of over 300,000 individuals, with information on the socio-economic and demographic characteristics of the respondent household and their location (home postcode). The survey response rate is around 1 completed survey for every 6 invites. The dataset comprises 1029 observations for

[8] England and Wales comprise a total of 1624 WSZ serving more than 50 million customers. The Drinking Water Inspectorate requires all WSZ to be monitored multiple times per year (see DWI, 2013). Sampling points within WSZ are randomly selected customer taps.

water hardness and 1074 for aesthetic quality.[9]

Summary statistics for both samples are reported in Table 1, starting with quality measures. For water hardness, subjective quality is measured on a 1 to 5 scale (from very soft to very hard) while objective quality is measured in mg CaCO3 per liter. For aesthetic quality the objective measure captures the number of complaints about taste and odor per thousand customers in a given WRZ, and to be consistent I combine subjective ratings for taste and odor measured on a 1 to 5 scale (from bad to excellent).[10] I will return to the relationship between subjective and objective quality measures in the analysis below.

Avertive expenditures represent a response to dissatisfaction with the quality of water, and follow-up questions identify which purchases of substitute products constitute an avertive behavior. This is especially important for aesthetic quality, as only 1 in 3 respondents who report purchases of substitute products do so as an avertive behavior. More specifically, a dislike of the taste and odor of tap water is provided by around 27 percent of the respondents as their main concern, and is thereby the most frequent motivation for undertaking substitute actions. The second most common motivation (25 percent) concerns the convenience of bottled water, and the re-use of bottles by 12 percent. Expenses by these respondents are not included in Table 1 (nor in subsequent econometrics analysis).[11] For respondents who report purchases of multiple substitute products expenditures are summed across products.

For water hardness, just over ten percent of respondents report positive avertive expenditures, the most common being those on water softener products for washing machines, dishwashers and kettles. Average avertive expenditures in the sample is around GBP14 per year,

[9] Note that respondents who reportedly carried out some avertive behavior but did not report avertive expenditures are excluded from the sample (rather than treating their expenditures as 'zeros'). These missing observations represent around two percent for water hardness and slightly more than 10 percent for aesthetic quality. As shown below, balance tests suggest no statistically significant differences on observables.

[10] In fact average rating of odor (3.78) is slightly higher than that of taste (3.58), although the two outcomes are strongly correlated (0.79, p-val<0.001). Results from regressions suggest that around 80% of marginal WTP estimates can be attributed to taste improvements, the rest to odor (see Lanz and Provins, 2014). Given this, for the present analysis which focuses on endogeneity, I treat both ratings the same.

[11] Concerning other aesthetic characteristics not included in the analysis, dislike of the appearance of tap water was stated to be a motivation by relatively few households (5%) and ranked lower in considerations than the temperature of tap water (7%).

Table 1: Summary statistics and balance tests

	Water hardness (N=1029)			Aesthetic quality (N=1074)			
	Mean	Min	Max	Mean	Min	Max	Diff.
Subjective quality	3.40	1	5	7.47	2	10	–
	(1.17)			(1.78)			
Objective quality	168.76	15.00	387.75	0.43	0	1.71	–
	(93.74)			(0.28)			
Avertive expenditures	14.18	0	785	28.06	0	464	–
	(64.60)			(63.10)			
INCOME	33.45	5	180	33.66	5	180	-0.20
('000 GBP)	(23.63)			(24.34)			(-0.20)
AGE	55.56	18	88	54.74	18	88	0.82
(year)	(13.77)			(13.85)			(1.36)
FEMALE	0.45	0	1	0.45	0	1	-0.01
(=1)	(0.50)			(0.50)			(-0.30)
HEALTHY	0.32	0	1	0.32	0	1	0.00
(=1)	(0.47)			(0.47)			(0.02)
HOSPITALIZED	0.11	0	1	0.11	0	1	0.00
(=1)	(0.31)			(0.31)			(-0.15)
RESIDENCY	14.20	0.5	25	13.95	0.5	25	0.24
(years)	(8.79)			(8.86)			(0.63)
HOME OWNER	0.48	0	1	0.47	0	1	0.01
(=1)	(0.50)			(0.50)			(0.45)
BILLS	0.36	0.1	0.65	0.36	0.1	0.65	0.00
('000 GBP)	(0.16)			(0.15)			(0.53)
FAMILY SIZE	2.25	1	7	2.26	1	7	0.00
(person)	(1.07)			(1.07)			(-0.07)
INFANTS	0.02	0	1	0.02	0	1	0.00
(infant)	(0.15)			(0.14)			(0.30)

Notes: Mean values are reported with standard deviations in parenthesis below. The column with "Diff." reports differences between samples means with t-statistics reported in parenthesis below. ***, **, *: statistically significant at 1, 5 and 10 percent respectively. Subjective quality for hardness are based on the scale 'very soft' (=1), 'soft' (=2), 'medium' (=3), 'hard' (=4), and 'very hard' (=5). Objective water hardness measured in mg CaCO3/l. Subjective quality for aesthetic quality is the sum of ratings for taste and odor of tap water based on the scale: 'bad' (=1); 'poor' (=2); 'adequate' (=3); 'good' (=4); and 'excellent' (=5). Objective aesthetic quality measured by the number of complaints per thousand customers in WSZ.

which is substantial if compared to average annual water bills of GBP180 (Ofwat, 2013).[12] Among those who reported positive avertive expenditures, the average is GBP114 and the median is GBP60. In relation to aesthetic quality, avertive behavior is more prevalent, as around 32 percent of respondents report positive avertive expenditures. The most common mitigating

[12] For simplicity capital expenditures are assumed to be equally spread over five years, so that avertive expenditures included in the analysis represents only one fifth of initial outlays plus any other recurring yearly expenditures. Other treatments of capital expenditures are of course possible, but implications for the results are minor and not particularly interesting.

behavior is the use of a jug with a filter (18.4%) followed by the purchase of bottled water (16.3%). Average expenditures are significantly higher than for water hardness (around GBP28). However, among respondents with positive expenditures, average expenditures are lower at GBP89, and median is GBP60. This indicates that, while the proportion of households with aesthetic quality related expenditures is greater than the proportion for water hardness, the individual amounts are smaller.

Finally, Table 1 provides summary statistics for socio-demographic variables included in the vector Z, together with t-tests for difference in means between subsamples. More specifically, covariates included are: household income in thousand GBP per year (INCOME); age in years (AGE); an indicator variable for gender (FEMALE, dummy); an indicator of weather the respondent assesses his health status to be better than that of someone with the same age (HEALTHY, dummy); an indicator of whether the respondent was hospitalized in the previous year (HOSPITALIZED, dummy); how long the respondent has been living in the same neighborhood, in years (RESIDENCY); whether she/he owns his home (HOME OWNER, dummy); yearly water bills in thousand GBP per year (BILLS); the number of family members in the household (FAMILY SIZE); the number of children below one year old in the household (INFANTS). None of the difference in means between samples are statistically significantly different from zero.

3.3 Estimation of the avertive expenditure function

Table 2 reports results for water hardness valuation function with subjective rating (specification I) and subjective rating instrumented with objective water hardness measured at the WSZ level (specification II). For tobit models both coefficient estimates and marginal effects evaluated at the mean of the sample are reported. Specification II also provides results from the first stage regression of subjective quality rating on objective quality and other controls, and displays the partial F-statistic associated with objective quality.

The Wald statistic for exogeneity, which provides evidence that ρ is different from zero (p-value <0.01), suggests that the perception of water hardness is endogenous. This confirms the expectation that unobserved heterogeneity affects both valuation of water hardness and its perception, and is consistent with Adamowicz et al. (2014) and Bontemps and Nauges (2015).

Table 2: Hardness of tap water – Household valuation function

	(I) Tobit model		(II) Tobit model with IV		
	β, γ	$\frac{\partial E(AVC\|AVC \geq 0, \overline{w}, Z)}{\partial(\overline{w}, Z)}$	First stage	β, γ	$\frac{\partial E(AVC\|AVC \geq 0, \overline{w}, Z)}{\partial(\overline{w}, Z)}$
RATING	77.51***	7.36***	–	133.25***	13.36***
	(14.97)	(1.29)		(24.47)	(2.59)
INCOME	1.28***	0.12***	0.001	0.98**	0.10**
	(0.44)	(0.04)	(0.001)	(0.45)	(0.05)
AGE	4.32***	0.41***	0.001	4.37***	0.44***
	(1.22)	(0.11)	(0.003)	(1.23)	(0.13)
FEMALE	-34.96	-3.28	0.073	-38.48	-3.81
	(25.25)	(2.30)	(0.057)	(25.68)	(2.35)
HEALTHY	11.02	1.06	-0.191***	25.59	2.65
	(27.01)	(2.65)	(0.060)	(28.39)	(3.01)
HOSPITALIZED	-12.51	-1.14	0.037	-17.55	-1.68
	(38.22)	(3.37)	(0.091)	(38.64)	(3.46)
RESIDENCY	2.32	0.22	0.010*	1.82	0.18
	(2.45)	(0.23)	(0.005)	(2.47)	(0.25)
HOME OWNER	121.54**	12.06**	0.072	121.09**	12.65**
	(59.23)	(5.99)	(0.122)	(59.46)	(6.20)
RESIDENCY x OWNER	-7.18**	-0.68**	-0.007	-6.86**	-0.69**
	(3.45)	(0.31)	(0.007)	(3.44)	(0.33)
BILLS	0.86	0.08	-0.032	48.81	4.89
	(80.91)	(7.68)	(0.185)	(80.58)	(7.98)
FAMILY SIZE	-3.03	-0.29	0.044	-4.11	-0.41
	(13.50)	(1.28)	(0.029)	(13.58)	(1.36)
INFANTS	8.17	0.80	-0.312*	26.54	2.91
	(84.23)	(8.46)	(0.187)	(81.24)	(9.71)
MG $CaCO3/l$	–	–	0.008***	–	–
			(0.000)		
CONSTANT	-871.51***		1.736***	-1069.90***	
	(140.74)		(0.181)	(167.50)	
σ^2	240.29***		–	235.64***	
	(26.82)			(25.75)	
Log-(pseudo)likelihood	-1088.76		–	-2397.08	
(Pseudo) R^2	0.038		0.445	–	
N	1029		1029	1029	
F-test/Wald (p-val)	0.00		0.00	0.00	
Wald stat. exo. (p-val)	–		–	0.00	
Partial 1^{st} stage F-stat.	–		749.7	–	

Notes: Table reports tobit estimates in specification (I) and simultaneous-equation tobit model in specification (II). Marginal effects evaluated at the sample mean. Robust standard errors are reported in parenthesis. ***, **, *: statistically significant at 1, 5 and 10 percent respectively.

Furthermore the first stage partial F-statistic for objective rating is well above values that would raise concerns about the validity (or weakness) of the instrument. To provide further confidence in the proposed instrument, I also implemented the minimum distance estimation approach for the structural model by Magnusson (2010), which allows for robust inference in the presence of potentially weak instruments.[13] Findings suggest that the weak-instrument robust 95 percent confidence interval for perceived quality [93.0, 174.2] is virtually identical to the usual Wald-type confidence interval [92.6, 173.9]. Comparing marginal effects reported in specification (I) and (II), the endogeneity bias is negative and economically relevant, as the marginal WTP evaluated at the mean of the sample increases by 86 percent.

Marginal effects for other covariates included in the analysis all have intuitive interpretations, and these are not significantly affected by using a simultaneous-equation setting. The effect of INCOME is positive and highly statistically significant, while AGE and HOME OWNERSHIP are positively related to avertive expenditures, and can be linked to the experience of the effect of water hardness on the lifetime of consumer appliances. Avertive expenditures are found to decline with how long the respondent has owned the home, which could indicate an adaptation to the local level of water hardness. As should be expected for hardness-related expenditures the health indicators, family indicator and other controls are not found to have a statistically impact on expenditures.

Results for avertive behavior in relation to the aesthetic quality of tap water are presented in Table 3. Specification (I) is again a standard tobit model with perceived quality rating and specification (II) uses the complaint rate to instrument for perceived quality rating.

For aesthetic quality, evidence about endogeneity of subjective rating is mixed. On the one hand, the Wald test for exogeneity is just out of the standard statistical confidence bounds (p-value=0.113). This is not completely unexpected because the potential for preference learning as a source of endogeneity is limited due to the availability of (perceived) high quality bottled water. Concerns about the complaint rate being a weak instrument also arise, as the partial F-statistic from the first stage regression (11.07) is near the cutoff values of Staiger and Stock (1997) and Stock and Yogo (2005). The potential for a weak instrument is confirmed

[13] This procedure, implemented in Stata by Finlay and Magnusson (2009), provides evidence about whether the confidence interval of the potentially endogenous variable remains stable when relaxing the assumption that the instruments are strong, albeit at the cost of assuming homoscedastic errors.

Table 3: Aesthetic quality of tap water – Household valuation function

	(I) Tobit model		(II) Tobit model with IV		
	β, γ	$\frac{\partial E(AVC \mid AVC \geq 0, \overline{w}, Z)}{\partial(\overline{w}, Z)}$	First stage	β, γ	$\frac{\partial E(AVC \mid AVC \geq 0, \overline{w}, Z)}{\partial(\overline{w}, Z)}$
RATING	-39.81***	-10.99***	–	-97.64***	-31.23***
	(3.28)	(0.83)		(36.79)	(11.66)
INCOME	0.43**	0.12**	0.004*	0.71**	0.23**
	(0.21)	(0.06)	(0.002)	(0.35)	(0.11)
AGE	0.42	0.12	0.003	0.54	0.17
	(0.54)	(0.15)	(0.005)	(0.64)	(0.20)
FEMALE	-38.32***	-10.43***	-0.106	-46.02***	-14.53***
	(10.88)	(2.97)	(0.105)	(13.42)	(4.15)
HEALTHY	7.07	1.98	0.349***	26.54	8.77
	(11.56)	(3.27)	(0.110)	(18.00)	(6.09)
HOSPITALIZED	26.04	7.90	-0.462***	-3.02	-0.96
	(16.51)	(5.46)	(0.164)	(27.86)	(8.77)
RESIDENCY	-0.51	-0.14	0.004	-0.20	-0.06
	(1.11)	(0.31)	(0.009)	(1.25)	(0.40)
HOME OWNER	31.24	8.72	0.169	57.47*	18.66*
	(24.09)	(6.81)	(0.224)	(32.40)	(10.64)
RESIDENCY x OWNER	-0.51	-0.14	-0.005	-1.64	-0.52
	(1.41)	(0.39)	(0.013)	(1.76)	(0.56)
BILLS	-5.91	-1.63	-0.024	-3.07	-0.98
	(34.35)	(9.48)	(0.340)	(40.02)	(12.80)
FAMILY SIZE	5.00	1.38	0.042	5.92	1.89
	(5.65)	(1.56)	(0.052)	(6.27)	(2.01)
INFANTS	-21.79	-5.44	-0.827*	-70.80	-17.81
	(46.11)	(10.36)	(0.433)	(69.04)	(13.05)
FEMALE x INFANTS	81.78	31.90	1.067	173.33*	89.98
	(56.83)	(28.97)	(0.657)	(94.21)	(67.00)
COMPLAINTS	–	–	-0.604***	–	–
			(0.182)		
CONSTANT	182.69***		7.237***	588.80**	
	(36.75)		(0.314)	(257.08)	
σ^2	129.02***		–	128.77***	
	(6.91)			(6.91)	
Log-(pseudo)likelihood	-2427.29		–	-4548.70	
(Pseudo) R^2	0.043		0.035	–	
N	1074		1074	1074	
F-test/Wald (p-val)	0.00		0.00	0.00	
Wald stat. exo. (p-val)	–		–	0.113	
Partial 1^{st} stage F-stat.	–		11.07	–	

Notes: Table reports tobit estimates in specification (I) and simultaneous-equation tobit model in specification (II). Marginal effects evaluated at the sample mean. Robust standard errors are reported in parenthesis. ***, **, *: statistically significant at 1, 5 and 10 percent respectively.

by evidence from the weak-instrument-robust 95 percent confidence interval of Magnusson (2010), which is [-276.8,-34.3], whereas the Wald confidence interval from the simultaneous-equation model is [-174.2,-21.1]. Hence while there is statistically robust evidence that the instrumented marginal WTP has the correct sign and is likely higher than that derived from a single-equation tobit model (which has a 95 percent confidence interval of [-46.2, -33.4]), the point estimate from the simultaneous-equation tobit model might be biased. Again this is not completely unexpected, since complaint rate in the respondent's WSZ is only an imperfect measure of objective aesthetic quality, and is itself derived from subjective measures (complaints).

3.4 Alternative specifications

The aim of this section is to provide evidence from two alternative specifications for the valuation function. The first deals with the selection of covariates included in the valuation function, which covers a number of potentially relevant household characteristics (income and water bills, health, household composition, and residency status), but remains nevertheless rather arbitrary. The second specification is concerned with the assumption of joint normality of outcome variables imposed by the simultaneous-equation tobit model.

To assess the impact of selecting covariates, consider the 'minimal' specification required to obtain marginal WTP estimates. Specifically I estimate a model in which the only covariate included is the rating variable. Table 4 reports the results for (I) a single-equation tobit model and (II) a simultaneous-equation tobit model that only includes the rating of water hardness as an explanatory variable.

Results indicate that marginal WTP estimates are largely unaffected by the choice of covariates. Specifically, in single-equation tobit models marginal WTP increases from 7.36 without covariates to 8.92 with the full list of covariates. More importantly, the simultaneous-equation correction remains statistically valid, and the coefficient of the first stage and results for the marginal WTP estimates barely change (decreasing from 13.36 with the full list of covariates to 13.21 in the model with no covariates).

Table 5 provides similar evidence for aesthetic quality. Marginal WTP estimates remain largely unaffected by the inclusion of covariates. In single-equation tobit model marginal WTP

Table 4: Hardness of tap water – Minimal valuation function

	(I) Tobit model		(II) Tobit model with IV		
	β, γ	$\frac{\partial E(AVC \mid AVC \geq 0, \overline{w}, Z)}{\partial(\overline{w}, Z)}$	First stage	β, γ	$\frac{\partial E(AVC \mid AVC \geq 0, \overline{w}, Z)}{\partial(\overline{w}, Z)}$
RATING	82.92***	8.61***	–	136.12***	14.65***
	(16.58)	(1.58)		(25.86)	(2.72)
MG $CaCO3/l$	–	–	0.008***	–	–
			(0.000)		
CONSTANT	-600.38***		2.002***	-784.63***	
	(91.17)		(0.056)	(123.89)	
σ^2	252.74***		–	248.65***	
	(29.78)			(29.02)	
Log-(pseudo) likelihood	-1107.23		–	-2426.26	
(Pseudo) R^2	0.022		0.438	–	
N	1029		1029	1029	
F-test/Wald (p-val)	0.00		0.00	0.00	
Wald stat. exo. (p-val)	–		–	0.00	
Partial 1^{st} stage F-stat.	–		801.4	–	

Notes: Table reports tobit estimates in specification (I) and simultaneous-equation tobit model in specification (II). Marginal effects evaluated at the sample mean. Robust standard errors are reported in parenthesis. ***, **, *: statistically significant at 1, 5 and 10 percent respectively.

declines from 10.99 with covariates to 10.90 without covariates, and in the simultaneous-equation tobit model it decreases from 31.23 with covariates to 28.79 without.

In the case of aesthetic quality, one important finding is that removing covariates provides sharper evidence about potential endogeneity of perceived quality and validity of the instrument. Indeed the p-value of the Wald test for exogeneity is now below the 10% threshold, confirming that the marginal WTP in the single-equation model is likely biased towards zero. Furthermore the partial F-statistic of the first stage is around 19, while the coefficient from the first stage equation remain close to those in Table 3. Bounds from the robust 95 percent confidence interval [-169.4,-48.9] are also much closer to those of the traditional Wald confidence interval [-143.9, -41.4]. Note that these confidence intervals both exclude the point estimate for marginal WTP obtained from the single-equation model. These results therefore support the view that marginal WTP estimates derived from the single-equation tobit are biased downward.

Table 5: Aesthetic quality of tap water – Minimal valuation function

	(I) Tobit model		(II) Tobit model with IV		
	β, γ	$\frac{\partial E(AVC \mid AVC \geq 0, \overline{w}, Z)}{\partial(\overline{w}, Z)}$	First stage	β, γ	$\frac{\partial E(AVC \mid AVC \geq 0, \overline{w}, Z)}{\partial(\overline{w}, Z)}$
RATING	-38.75***	-10.86***	–	-108.21***	-36.25***
	(3.33)	(0.85)		(41.64)	(13.85)
INQUIRIES	–	–	-0.637***	–	–
			(0.145)		
CONSTANT	212.09***		7.616***	730.77**	
	(21.40)		(0.075)	(310.11)	
σ^2	132.91***		–	132.56***	
	(7.40)			(7.40)	
Log-(pseudo) likelihood	-2445.27		–	-4583.71	
(Pseudo) R^2	0.036		0.01	–	
N	1074		1074	1074	
F-test/Wald (p-val)	0.00		0.00	0.00	
Wald stat. exo. (p-val)	–		–	0.094	
Partial 1^{st} stage F-stat.	–		19.39	–	

Notes: Table reports tobit estimates in specification (I) and simultaneous-equation tobit model in specification (II). Marginal effects evaluated at the sample mean. Robust standard errors are reported in parenthesis. ***, **, *: statistically significant at 1, 5 and 10 percent respectively.

A second source of concerns with the simultaneous-equation tobit model is related to the joint normality assumption. As an alternative, a linear two stage least square (2SLS) specification is employed. This provides consistent estimates of partial effects although it does not account for the truncated nature of the avertive expenditure data.

Results reported in Table 6 show that for both hardness and aesthetic quality marginal WTP estimates are very close to those derived from the simultaneous-equation tobit model. Specifically, for water hardness marginal WTP estimated from a 2SLS model is 13.78, and 13.36 from the simultaneous-equation tobit model (see Table 2). For aesthetic quality, marginal WTP from the 2SLS model is -27.95, which is slightly lower than that derived from the simultaneous-equation tobit model, -31.23 (Table 3). Consistency in the results from different estimation strategies suggest that the distributional assumptions underlying the simultaneous-equation tobit models do not significantly affect the results.

Taken together, evidence from alternative specifications provide confidence in the marginal WTP estimates derived from the simultaneous-equation tobit model. This is especially the

Table 6: Linear 2SLS estimation – Household valuation function

	(I) Water hardness	(II) Aesthetic quality
RATING	13.78***	-27.95**
	(2.94)	(12.38)
CONSTANT	-72.24***	211.79**
	(17.49)	(86.67)
R^2	0.085	0.100
N	1029	1074
F-test (p-val)	0.00	0.00
K-P F-stat.	789.40	7.38

Notes: Table reports 2SLS estimates for water hardness in specification (I) and aesthetic quality in specification (II). The following covariates are included: INCOME, AGE, FEMALE, HEALTHY, HOSPITALIZED, RESIDENCY, HOME, RESIDENCY X OWNER, BILLS, HOUSEHOLD, INFANTS, and for the aesthetic quality equation FEMALE X INFANTS. Weak identification is informed by the Kleibergen-Paap (K-P) F-statistic. Robust standard errors are reported in parenthesis. ***, **, *: statistically significant at 1, 5 and 10 percent respectively.

case for water hardness, for which results are encouraging. For aesthetic quality, the measure of objective quality is not as good, and although evidence derived alternative specifications suggests that endogeneity of perceived quality is likely an issue, results are potentially subject to a weak instrument bias.

3.5 Marginal willingness to pay and objective quality

As mentioned above marginal effects are highly non-linear functions of the estimated parameters. Based on simultaneous-equation tobit specification reported in Table 2 and 3, Table 7 provides further evidence about how marginal WTP for water softening and aesthetic quality vary across income and ratings.

Further, results from the first stage regressions can be used to relate marginal WTP estimates to changes in objective quality. Typically, objective quality is the outcome targeted by investments in infrastructures, whereas perceived quality would be affected only indirectly. For water hardness, the highest concentration measured is around 400 mg CaCO3 per liter and reducing this to around 60 would correspond to a change in subjective rating by $340 \cdot 0.008 = 2.72$, or roughly GBP35 per household per year. For aesthetic quality a reduction in the number of complaints from is maximum level of 0.003 per thousand of customer

Table 7: Marginal willingness to pay estimates – water softening and aesthetic quality

		Average avertive expenditure[a] $E(Y\|Y \geq 0, X)$			mWTP (GBP/household/year) $\frac{\partial E(AVC\|AVC \geq 0, \overline{w}, Z)}{\partial \overline{w}}$		
Yearly income ('000GBP)		25	40	65	25	40	65
Hardness rating	2	1.77***	2.08***	2.69***	2.60*	2.99**	3.77**
	3	6.82***	7.81***	9.71***	8.43***	9.45***	11.36***
	4	21.03***	23.5***	28.11***	21.43***	23.40***	26.95***
Aesthetic rating	3	354.21***	364.73***	382.30***	-96.11**	-96.34**	-96.66**
	5	172.69***	181.65***	196.94***	-81.25*	-82.80*	-85.15*
	8	5.79***	6.68***	8.41***	-7.64***	-8.61***	-10.43***

Notes: Average avertive expenditures and marginal WTP in GBP per household per year are evaluated at the mean of the sample. ***, **, *: statistically significant at 1, 5 and 10 percent respectively based on robust standard errors.

in a WSZ to zero corresponds to an increase of quality rating by $0.003 \cdot 1000 \cdot -0.604 \simeq 1.8$. Evaluated towards the low end of the rating scale reported in Table 7, this is equivalent to a monetary value of more than GBP170 per household per year.

4 Discussion and conclusion

This paper has studied the issue of endogenous quality perception in estimating marginal WTP for water quality improvements. While variations in objective quality are often employed to identify the demand schedule for public goods and used to inform the socially optimal provision, objective quality only indirectly determines choices by households. However, using perceived quality in the valuation function instead may lead to an endogeneity bias when some variables affect both perception and avertive expenditures.

The main contribution of this paper has been to propose the use an objective quality measure as an instrument to recover WTP estimates for marginal changes in perceived quality. On the one hand, I motivated the exclusion restriction with a simple theoretical framework, as objective quality can be expected to affect valuation only through perceived quality. On the other hand, the approach permits estimating marginal WTP estimates in relation to objective quality, as the first stage regression quantifies the relationship between perceived and objective quality. This allows obtaining micro-consistent estimates of marginal WTP for improvements in the objective provision of the public good, which is the more relevant measure from a policy perspective.

As an illustration of the proposed approach, I have used data on avertive expenditures for water softening devices and substitutes to aesthetic quality of tap water. First, I have shown that households are actively responding to variation in service levels concerning tap water quality, even when these do not concern health-related risks. While actual observed expenditure on substitute products are relatively minor in terms of overall household budget, it is proportionately large in comparison to the average water services bill. Second, evidence suggests that perceived quality is endogenous and the associated marginal WTP estimates are biased towards zero. In the setting considered, this could be due to a learning process about the benefits of substitute products, but it could also be related to other unobserved households characteristics affecting both ratings and expenditures.

The crucial element for the approach to be applicable is, of course, that a first stage actually exists. In other words, instrumenting perceived quality with objective quality can only work in settings for which (i) objective quality is exogenous and observed by consumers; (ii) consumers use this information to shape their subjective quality perception; and (iii) the analyst can obtain a measure of both perceived and objective quality that is relevant from a decision-maker's perspective. Clearly, for some goods these requirements cannot be met. In the case of aesthetic quality, I had to rely on regional complaint rates as a measure of objective quality, which raised some concerns about the results. For water hardness, where a good physical measure of objective quality was available, results suggested that the proposed instrumental variable strategy works quite well. Therefore, while it is important to stress that the specific marginal WTP estimates reported here rely on a number of assumptions that should be further scrutinized (such as the treatment of capital expenditures), the proposed approach can potentially be applied to other settings.

References

Abdalla, C., B. Roach, and D. Epp (1992) "Valuing environmental quality changes using averting expenditures: An application to groundwater contamination," *Land Economics*, 68 (2), pp. 163 – 169.

Abrahams, N., B. Hubbell, and J. Jordan (2000) "Joint production and averting expenditure measures of willingness to pay: Do water expenditures really measure avoidance costs?" *American Journal of Agricultural Economics*, 82(2), pp. 427 – 437.

Adamowicz, W., D. Dupont, P. Lloyd-Smith, and C. Schram (2014) "Endogeneity of risk perceptions in water expenditure models." Working paper.

Amemiya, T. (1979) "The estimation of a simultaneous-equation tobit model," *International Economic Review*, 20 (1), pp. 169 – 181.

Bartik, T. (1988) "Evaluating the benefits of non-marginal reductions in pollution using information on defensive expenditures," *Journal of Environmental Economics and Managment*, 15, pp. 111 – 127.

Bayer, P., N. Keohane, and C. Timmins (2009) "Migration and hedonic valuation: The case of air quality," *Journal of Environmental Economics and Management*, 58, pp. 1–14.

Becker, G. S. (1965) "A theory of the allocation of time," *Economic Journal*, 75, pp. 493 – 517.

Bontemps, C. and C. Nauges (2009) "Carafe ou bouteille? Le rôle de la qualité de l'environnement dans la décision du consommateur," *Economie et Prévision*, 188, pp. 61 – 79.

——— (2015) "The impact of perceptions in averting-decision models: An application of the special regressor method to drinking water choices," *American Journal of Agricultural Economics*, forthcoming.

Cameron, T. (1988) "A new paradigm for valuing non-market goods using referendum data: Maximum likelihood estimation by censored logistic regression," *Journal of Environmental Economics and Management*, 15, pp. 355 – 379.

Cameron, T. and J. Englin (1997) "Respondent experience and contingent valuation of environmental goods," *Journal of Environmental Economics and Management*, 33 (3), pp. 296 – 313.

Cameron, T. and M. James (1987) "Efficient estimation methods for contingent valuation surveys," *Review of Economics and Statistics*, 69, pp. 269 – 276.

Courant, P. and R. Porter (1981) "Averting expenditure and the cost of pollution," *Journal of Environmental Economics and Management*, 8 (4), pp. 321 – 329.

Czajkowski, M., N. Hanley, and J. LaRiviere (2014) "The effect of experience on preferences: Theory and empirics for environmental public goods," *American Journal of Agricultural Economics*, 97, pp. 333 – 351.

Danielson, L., T. Hoban, G. VanHoutven, and J. Whitehead (1995) "Measuring the benefits of local public goods: Environmental quality in Gaston County, North Carolina," *Applied Economics*, 27, pp. 1253 – 1260.

Deschenes, O., M. Greenstone, and J. Shapiro (2012) "Defensive investments and the demand for air quality: Evidence from the NO_x budget program and ozone reductions." MIT CEEPR working paper.

Dickie, M. and S. Gerking (1996) "Formation of risk beliefs, joint production and willingness to pay to avoid skin cancer," *The Review of Economics and Statistics*, 78 (3), pp. 451 – 463.

Dupont, D. (2005) "Tapping into consumers' perception of drinking water quality in Canada: Capturing customer demand to assist in better management of water resources," *Canadian Water Resource Journal*, 30 (1), pp. 11 – 20.

Dupont, D. and N. Jahan (2012) "Defensive spending on tap water substitutes: The value of reducing perceived health risks," *Journal of Water and Health*, 10(1), pp. 56 – 68.

DWI (2013) "Drinking water 2012; A report by the Chief Inspector of drinking water." Drinking Water Inspectorate, UK.

Finlay, K. and L. Magnusson (2009) "Implementing weak-instrument robust tests for a general class of instrumental-variables models," *The Stata Journal*, 9 (3), pp. 398 – 421.

Gerking, S. and L. Stanley (1986) "An economic analysis of air pollution and health: The case of St. Louis," *Review of Economics and Statistics*, 68, pp. 115 – 121.

Grossman, M. (1972) "On the concept of health capital and the demand for health," *Journal of Political Economy*, 80 (2), pp. 223 – 255.

Harford, J. (1984) "Averting behavior and the benefits of reduced soiling," *Journal of Environmental Economics and Management*, 11, pp. 296 – 302.

Harrington, W., A. Krupnick, and W. Spofford (1989) "The economic losses of a waterborne disease outbreak," *Journal of Urban Economics*, 25, pp. 116 – 137.

Harrington, W. and P. Portney (1987) "Valuing the benefits of health and safety regulation," *Journal of Urban Economics*, 22, pp. 101 – 112.

Jakus, P. M., W. Shaw, N. Nguyen, and M. Walker (2009) "Risk perceptions of arsenic in tap water and consumption of bottled water," *Water Resources Research*, 45 (5), p. W05405.

Lanz, B. and A. Provins (2014) "Estimating the demand for tap water quality: Avertive expenditures on substitutes for hardness and aesthetic quality." CIES working paper.

Larson, B. and E. Gnedenko (1999) "Avoiding health risks from drinking water in Moscow: An empirical analysis," *Environment and Development Economics*, 4, pp. 565 – 581.

Magnusson, L. M. (2010) "Inference in limited dependent variable models robust to weak identification," *Econometrics Journal*, 13 (3), pp. S56–S79.

McConnell, K. and M. Rosado (2000) "Valuing discrete improvements in drinking water quality through revealed preferences," *Water Resources Research*, 36, pp. 1575 – 1582.

Nauges, C. and C. van den Berg (2009) "Perception of health risk and averting behavior: The case of water treatment in Southwest Sri Lanka." Toulouse School of Economics working paper 09-139.

Nelson, F. and L. Olson (1978) "Specification and estimation of a simultaneous-equation model with limited dependent variables," *International Economic Review*, 19 (3), pp. 695 – 709.

Ofwat (2013) "Average household bills fact sheet." see http://www.ofwat.gov.uk/regulating/charges/prs_web_charges2013-14 (accessed January 2014).

Orgill, J., A. Jeuland, J. Brown, and A. Shaheed (2013) "Water quality perceptions and willingness to pay for clean water in peri-urban cambodian communities," *Journal of Water and Health*, 11, pp. 489 – 506.

Rosado, M., M. Cunha-E-Sa, M. Ducla-Soares, and L. Nunes (2006) "Combining averting behavior and contingent valuation data: An application to drinking water treatment in Brazil," *Environment and Development Economics*, 11 (6), pp. 729 – 746.

Schram, C., W. Adamowicz, and D. Dupont (2010) "At odds with water: Percieved health risks and water in Canada." Working paper.

Slovic, P. (2000) *The perception of risk*: London: Earthscan.

Smith, R. and R. Blundell (1986) "An exogeneity test for a simultaneous equation tobit model with an application to labor supply," *Econometrica*, 54, pp. 679 – 685.

Smith, V. and W. Desvousges (1986) "Averting behavior: Does it exist?" *Economics Letters*, 20, pp. 291 – 296.

Staiger, D. and J. Stock (1997) "Instrumental variables regression with weak instruments," *Econometrica*, 65 (3), pp. 557 – 586.

Stock, J. and M. Yogo (2005) "Testing for weak instruments in linear iv regression," in D. Andrews and J. Stock eds. *Identifcation and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*: New York: Cambridge University Press.

Tra, C. (2010) "A discrete choice equilibrium approach to valuing large environmental changes," *Journal of Public Economics*, 94 (1-2), pp. 183 – 196.

Um, M.-J., S.-J. Kwak, and T.-Y. Kim (2002) "Estimating willingness to pay for improved drinking water quality using averting behavior method with perception measure," *Environmental and Resource Economics*, 21, pp. 287 – 302.

Whitehead, J. . (2006) “Improving willingness to pay estimates for quality improvements through joint estimation with quality perceptions,” *Southern Economic Journal*, 73 (1), pp. 100 – 111.

Whitehead, J., G. Blomquist, T. Hoban, and W. Clifford (1995) “Assessing the validity and reliability of contingent values: A comparison of on-site users, off-site users, and non-users,” *Journal of Environmental Economics and Management*, 29 (2), pp. 238 – 251.

Yoo, S.-H. and C.-Y. Yang (2000) “Dealing with bottled water expenditures data with zero observations: A semiparametric specification,” *Economics Letters*, 66, pp. 151 – 157.

Zivin, J. G., M. Neidell, and W. Schlenker (2011) “Water quality violations and avoidance behavior: Evidence from bottled water consumption,” *American Economic Review*, 101(3), pp. 448 – 453.