University of Neuchatel

Institute of Economic Research

IRENE Working paper 16-06

Using averting expenditures to estimate the demand for public goods: Combining objective and perceived quality

Bruno Lanz

Allan Provins

Institut de recherches économiques

Using averting expenditures to estimate the demand for public goods: Combining objective and perceived quality*

Bruno Lanz† Allan Provins‡

This version: October 2016

Abstract

In response to the perceived quality of a public good, households may choose to incur averting expenditures as a substitute to its aggregate provision, thereby revealing an (inverse) demand function. When unobserved heterogeneity affects both perceived quality and averting behavior, identification of the demand function is plagued by a problem of endogeneity. In this paper, we propose the use of an auxiliary (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 averting expenditures method, and when objective quality of provision is plausibly exogenous. We illustrate the approach with data on averting 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; Averting expenditures; Revealed preferences; Perceived quality; Objective quality; Water demand.

JEL Codes: D1, H4, Q2, Q5.

---

*We would like to thank Ian Bateman, Diane Dupont, Jean-Philippe Nicolai, Scott Reid, and participants at EPFL and 2014 WCERE meeting for discussions and suggestions. We are grateful to two anonymous referees for comments that have substantially improved the paper. Excellent research assistance was provided by Martin Baker and Lawrie Harper-Simmonds. We also thank the participating water companies for allowing us to access their data and for partially funding this work. The views expressed in this paper and any remaining errors are ours alone.

†Corresponding author. University of Neuchâtel, Department of Economics and Business, Switzerland; Chair for Integrative Risk Management and Economics, ETH Zurich, Switzerland; Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, USA. Mail: A.-L. Breguet 2, CH-2000 Neuchâtel, Switzerland; Tel: +41 32 718 14 55; email: bruno.lanz@unine.ch.

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

1 Introduction

There is an established practice of using household's averting 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 averting 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 for applying the averting 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 averting behavior and expenditures (Dickie and Gerking, 1996; Abrahams et al., 2000). In instances where individuals have heterogeneous perception of the public good provision, it is *a priori* important to control for the relationship between objective provision and perceptions (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; see in particular Poe et al. (1998) and Poe and Bishop (1999) for empirical evidence on the impact of information on subjective risk perception and valuation. To account for such differences, a number of averting behavior studies have employed a measures of perceived provision to identify marginal WTP estimates (e.g. 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 averting behavior and quality perception, identification of marginal WTP with variations in perceived quality is likely to generate biased estimates (Whitehead, 2006; Nauges and van den Berg, 2009; Orgill et al., 2013; Adamowicz et al., 2014; Bontemps and Nauges, 2016).

In an attempt to address these two issues, this paper proposes to combine information on perceived (subjective) and objective provision level. We control for the potential endogeneity of perceived quality in the valuation function with a simultaneous equation estimation procedure, modeling the relationship between objective quality and subjective perception in an auxiliary (first stage) regression. 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, addressing the potential endogeneity of perceived quality. On the other hand, the auxiliary regression quantifies the relationship between subjective and objective provision, and can be used to obtain theoretically valid marginal WTP estimates in relation to the policy-relevant objective level of provision.[1] Importantly, the exclusion restriction relies on an assumption that objective quality is exogenous and affects the demand for marketed products only through perceived quality, and in the paper we emphasize conditions under which this is likely to be plausible.

To illustrate our approach, we employ data from a survey administered in England and Wa-

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

les eliciting averting expenditures for two characteristics of public tap water supply,[2] namely water hardness and the aesthetic quality in terms of taste and odor.[3] As we discuss further in the paper, for these qualitative aspects of tap water supply the validity of the proposed exclusion restriction is plausible. First, preference-based sorting (e.g. Chay and Greenstone, 2005; Bayer et al., 2009) is unlikely, as these aspects of water quality can safely be assumed not to enter individual's location decisions. Second, the water industry in England and Wales is composed of state-regulated regional monopolies, with strict compliance with European drinking water standards (The Drinking Water Directive, 98/83/EC), which makes the provision of these services unrelated to socio-economic outcomes. Therefore, in the particular setting we consider, variation in the objective level of hardness and aesthetic quality can plausibly be seen as exogenous, and can be used to identity marginal WTP estimates for improvements to these aspects of tap water supply.

Aside from recording averting 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 the consumption of a public good affects its valuation (e.g. Whitehead et al., 1995; Cameron and Englin, 1997; Czajkowski et al., 2014), and in the case of averting behavior products purchased on the market will provide consumers with an alternative experience of the good. This experience will then likely affect both perceived quality and averting expenditures. 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.[4]

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).[5] Specifically,

[2] In a companion paper, Lanz and Provins (2016), we provide a comprehensive description of wider survey results, including detailed evidence about the sort of market product purchased. In the present paper, we rather focus on the potential endogeneity of perceived quality and implications for welfare estimates.

[3] Hard water can reduce the lifetime of water-using appliances and thus impose financial costs on households.

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

[5] WSZ are geographical 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.

the measures of objective quality 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 80% higher for water hardness and almost three time higher for aesthetic quality.

This paper is closely related to two recent studies tackling the issue of endogenous quality perception in the context of an averting behavior model.[6] 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% lower as compared to instrumented estimates. Second, Bontemps and Nauges (2016) 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. In their work, the perception variable is instrumented 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 about general water pollution in that region.[7] 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%. Relative to the instrumental variable strategies used in these papers, the one proposed here is theoretically motivated, as the auxiliary regression 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 presents the analytical framework supporting the estimation. Section 3 describes the empirical illustration. Section 4 concludes.

[6] 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.

[7] 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 presents the framework supporting the econometric estimation of an averting expenditure function. We start with a standard averting expenditure model using the household production function framework of Becker (1965), Grossman (1972) and Courant and Porter (1981). Our exposition draws from Bartik (1988) who links averting 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). We 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 Averting 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 or use, 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 averting cost function *AVC*(q, w), with ∂*AVC*$/\partial q > 0$ and ∂*AVC*$/\partial w < 0$. This function summarizes expenditure needed to reach quality q given the specific provision level w (see e.g. Bartik, 1988; Um et al., 2002).[8] Given income M and budget constraint $M = X + AVC(q, w)$, consumers will purchase marketed products improving experienced quality until the desired 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), averting expenditures provide information about the value that consumers derive from 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 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:

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

[8] Using a reduced-form *AVC* function implicitly assumes that there is a continuum of possible improvements that translate into a smooth cost function. While this provides a useful specification to analyze marginal conditions, it is clearly a simplification, as consumers would typically make discrete purchase decisions. We note, however, that in our empirical application there are products available that are both low cost and improve quality at the margin, such as softening tablets for hardness and bottled water for aesthetic quality. Therefore the fact that consumers have access to these products “convexifies” the choice problem, as it essentially allows them to reach any desired level of (average) water quality, which is in line with the notion of a smooth AVC.

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 averting cost function, q^* will generally also change as $\overline{w}$ (and w) increase. Under the conditions discussed in Bartik (1988), variations in the averting cost function will 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) we 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 averting cost function (Cameron and James, 1987; Cameron, 1988). Hence for a consumer i we have that observed averting 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 averting 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:[9]

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

It follows that when $\rho = corr(\varepsilon_i, \eta_i) \neq 0$, the 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. Averting expenditures and quality perception are modeled as simultaneous processes, whereby changes in an objective measure of provision translates into an improved perceived quality, and improved perceived quality reduces averting 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, as in Tiebout's (1956) sorting mechanism. Intuitively, households with relatively strong preferences for some local public good may chose to locate in neighborhoods with relatively high provision of that good, and simultaneously undertake relatively more averting actions than a representative household would in the same neighborhood. Conversely, in areas with relatively low provision levels, individuals with relatively low preferences for that particular good would be over-represented, and we would observe lower averting expenditure compared to the case where the public good would not enter individuals' location decisions. Therefore, objectively measured quality of provision ($\overline{w}$) would affect both subjective rating and averting expenditures.

Examples of preference-based sorting include Chay and Greenstone (2005), Bayer et al. (2009) and Tra (2010), who provide evidence that households sort across neighborhoods

[9] Note that for simplicity we write perceived quality as a function of contemporaneous quality, although individual characteristics included in the vector S could include information about past experiences. The specification of the perception function is an empirical matter and would mainly depend on the specific goods considered and data availability.

based on prevailing air quality. For less salient amenities such as water quality, the exclusion restriction may also be violated if standards applied in certain areas differ systematically with socio-economic characteristics.[10] For example, in the U.S. Zivin et al. (2011) provide evidence that relatively poor neighborhoods face a higher risk of water quality violations. In turn, endogeneity of objective provision further complicates identification of marginal WTP values through an averting behavior approach (see Shogren and Crocker, 1991). However, in the context which we consider, namely the choice of qualitative aspects of tap water, this is unlikely to be an issue; we come back to the plausibility of our exclusion restriction in Section 3.4 below.

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. We treat averting expenditures as a corner solution outcome, assuming that households with zero expenditures optimally chose this amount, and we use a tobit model to represent the conditional expectation of expenditures. Observed averting 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.

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

[10] Similarly, for water quality choices, the exclusion restriction might breakdown 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 water supply networks, where source choice does not enter, this is less of an issue.

perceived quality are jointly estimated as:

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

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 averting expenditures conditional on the vector of covariates (w_i, Z_i), denoted by $E(\mathit{AVC}_i \,|\, \mathit{AVC}_i \geq 0, w_i, Z_i)$, can be decomposed into two parts:

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

Marginal WTP estimates for improved quality perception, $\frac{\partial E(\mathit{AVC}_i \,|\, \mathit{AVC}_i \geq 0, w_i, Z_i)}{\partial w_i}$, comprise both changes in the decision of whether or not to incur averting expenditures (or a change in the fraction of households with positive averting expenditures), as measured by $P(\mathit{AVC}_i > 0 \,|\, w_i, Z_i)$, and changes in the average amount spent by households who decide to incur expenditure, denoted $E(\mathit{AVC}_i \,|\, \mathit{AVC}_i > 0, w_i, Z_i)$. Marginal WTP estimates, and more generally marginal effects, are a highly non-linear functions 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:

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

where $\frac{\partial w_i}{\partial \overline{w}_i} = \phi$ is simultaneously estimated from the auxiliary (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, we illustrate the use of the estimation strategy in the context of tap water quality choices. More specifically, we consider two qualitative aspects of household tap water supply from water services utilities: water hardness and aesthetic quality in terms of taste and odor. Empirical results reported in this paper are based on a survey of customers of water companies in England and Wales (see Lanz and Provins, 2016). In the following, we first describe our survey instrument, data collection, and provide summary statistics. We then turn to the econometric estimation of the valuation function, discuss the validity of the exclusion restriction, and check that our results are robust to alternative econometric specifications. Finally results for marginal WTP estimates for perceived and objective quality changes are presented.

3.1 Survey instrument and data sources

The survey instrument was developed through several stages, including a national omnibus survey to determine the set of averting behaviors carried out by households and an online pilot with a sample of approximately 200 respondents. The final survey is structured as follows. After 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 market products improving tap water quality is then elicited, including water filters, bottled water, squash and cordial, water softener devices, and other products (e.g. tablets, powders and coils).

Based on the respondent's reported use of quality-improving products, they are presented with a series of follow-up questions in which they indicate the specific product types they use/purchase, their 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 incur expenditures related to tap water quality, notably reasons related to the aesthetic quality of tap water, water hardness and other motivations such as convenience or health concerns. This allows determining which expenditures are related to averting behavior.

The survey then asks respondents about specific experiences and perceptions related to the use of marketed products improving perceived water quality, including experience of problems with aesthetic quality of tap water, the quality of tap water (e.g. hardness, impurities) and health issues. 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 and waste water bill amount, their own health status and the health status of others in their household.

Data measuring objective quality of tap water is taken from mandatory reporting requirements by water companies to the Drinking Water Inspectorate for England and Wales. Objective quality measures relates to individual WSZ, which are geographical areas with an average population of around 30,000 supplied from the same source of treated water, and refers to the calendar year preceding the survey.[11] In each WSZ, an objective measure of water hardness is given by the average mg of calcium carbonate (CaCO3) per liter. For aesthetic quality, we 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. Survey 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 through an online platform, and our sample is drawn from a nation-wide sample of more than 300,000 individuals in England and Wales. Members of the panel are not allowed to opt into the survey, rather potential participants are randomly selected and invited to participate via an email message "We are conducting a survey on water services and tap water. If you would like to take part please click this link." Around one in six invitees actually completed the survey. The sampling process within the panel is further constrained by quotas set for socio-demographic characteristics (age, gender, socio-economic groups, and main newspaper read), so that by construction our sample is broadly representative of the underlying population (recall that we also keep only respondents either responsible or jointly responsible for paying household bills). Importantly, the generality of the invitation message should minimize the probability that respondents who

[11] 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.

experienced specific problems with water hardness or aesthetic quality self-selected into our sample.[12] The final data comprises 1029 observations for water hardness and 1074 for aesthetic quality, with summary statistics for both samples reported in Table 1.[13] Table 1 also reports average population characteristics for England and Wales taken from the 2011 census (Office for National Statistics, 2011), and household's average water and waste water bills (Ofwat, 2013).

Starting with quality measures, the subjective quality for water hardness 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 WSZ, and to be consistent we combine subjective ratings for taste and odor measured on a 1 to 5 scale (from bad to excellent).[14] We return to the relationship between subjective and objective quality measures in the analysis below.

Averting expenditures represent a response to dissatisfaction with the quality of water, and follow-up questions identify which purchases of marketed products constitute an averting behavior; see Lanz and Provins (2016) for more detailed evidence on product-specific expenditures. For aesthetic quality, a dislike of the taste and odor of tap water is provided by around 27% of the respondents as their main concern, and is thereby the most frequent motivation for undertaking averting actions. The second most common motivation (25%) concerns the convenience of bottled water, and the re-use of bottles by 12%. Expenses by these respondents are not included in Table 1 (nor in subsequent econometrics analysis). For respondents who

[12] While we cannot completely rule out the possibility that we disproportionately recruited individuals who experienced *general* problems with water services, we note that the effect of potential selection bias on our marginal WTP estimates is likely to be minimal. First, because we do not mention water hardness or aesthetic quality in our invitation message, any bias in averting expenditures for the dimensions of water services we consider would be small. Second, even if a small share of self-selected respondents could raise average averting expenditures, their presence is unlikely to significantly affect the relationship between the level of provision and expenditures (i.e. marginal WTP estimates). In particular, if respondents with low ratings also have higher expenditures, our marginal WTP estimates would rightly pick up how relatively low perception of hardness / aesthetic quality translates into higher average expenditures.

[13] Note that respondents who reportedly carried out some averting behavior but did not report averting 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% for aesthetic quality.

[14] More specifically, we sum rating variables for taste and odor. As an alternative, we have also carried out a factor analysis with the two rating variables, although we found that the resulting latent index is almost perfectly correlated with the sum of rating variables. We therefore employ our simpler and more transparent sum of ratings as our measure of subjective rating.

Table 1: Summary statistics

	Water hardness (N=1029)			Aesthetic quality (N=1074)			England and
	Mean	Min	Max	Mean	Min	Max	Wales[a]
Subjective quality	3.40 (0.04)	1	5	7.47 (0.05)	2	10	–
Objective quality	168.76 (2.92)	15.00	387.75	0.43 (0.01)	0	1.71	–
Averting expenditures (GBP/household/year)	14.18 (2.01)	0	785	28.06 (1.93)	0	464	–
Positive averting expenditures[a]	113.98 (13.20)	1.5	785	88.91 (4.61)	3.5	464	–
INCOME ('000 GBP)	30.45 (0.74)	5	180	31.66 (0.74)	5	180	32.11
AGE (year)	55.56 (0.43)	21	88	54.74 (0.42)	21	88	47.36
FEMALE (=1)	0.45 (0.02)	0	1	0.45 (0.02)	0	1	0.51
HEALTHY (=1)	0.32 (0.01)	0	1	0.32 (0.01)	0	1	–
HOSPITALIZED (=1)	0.11 (0.01)	0	1	0.11 (0.01)	0	1	–
RESIDENCY (years)	14.20 (0.27)	0.5	25	13.95 (0.27)	0.5	25	–
HOME OWNER (=1)	0.48 (0.02)	0	1	0.47 (0.02)	0	1	0.64
BILLS ('000 GBP)	0.36 (0.001)	0.1	0.65	0.36 (0.001)	0.1	0.65	0.38
FAMILY SIZE (# person)	2.25 (0.03)	1	7	2.26 (0.03)	1	7	2.41
INFANTS (# infant)	0.02 (0.001)	0	1	0.02 (0.001)	0	1	–

Notes: Mean values are reported with standard error in parenthesis below. 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); see footnote 14. Objective aesthetic quality measured by the number of complaints per thousand customers in WSZ. [a] Summary expenditures for respondents who reported positive spending; units are GBP/household/year. The percentage of respondents who reported zero expenditures is 87.6 for water hardness and 68.4 for aesthetic quality. [b] Data for England and Wales taken from the 2011 census (Office for National Statistics, 2011), except for household's water and waste water bills (Ofwat, 2013).

report purchases of multiple products expenditures are summed across products.[15]

For water hardness, around twelve percent of respondents report positive averting expenditures, the most common being those for water softener products for washing machines, dishwashers and kettles. Average averting expenditures in the sample is around GBP 14 per year, and among those respondents who reported positive averting expenditures the average is around GBP 114. In relation to aesthetic quality, averting behavior is more prevalent, as around 32% of respondents report positive averting expenditures. Average expenditures are significantly higher than for water hardness (around GBP 28), but among respondents with positive expenditures average expenditures are lower than for water hardness (around GBP 89). The most common mitigating behavior for aesthetic quality is the use of a jug with a filter followed by the purchase of bottled water.

Finally, Table 1 provides summary statistics for socio-demographic variables included in the vector Z, namely: 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 and waste water bills in thousand GBP per year (BILLS); the number of family members in the household (FAMILY SIZE); the number of children below two years of age in the household (INFANTS). As compared to population-level statistics for a number of key variables, we find a number of small discrepancies, the most important one being that our sample is slightly older than the population and that it also includes a lower share of home owners. For the empirical analysis below, we therefore use population-level statistics to reweigh our results to be representative of the England and Wales population.

[15] For simplicity capital expenditures are assumed to be equally spread over five years, so that averting 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.

3.3 Estimation of the averting expenditure function

Table 2 reports results for water hardness valuation function with subjective rating (tobit model, specification I) and subjective rating instrumented with objective water hardness measured at the WSZ level (simultaneous equation tobit model, specification II). We report both coefficient estimates and marginal effects evaluated at the mean of the sample for the tobit model results. Specification II also provides results from the auxiliary (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 (2016). 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, we 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.[16] Findings suggest that the weak-instrument robust 95% 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 83%.

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 averting expenditures, and can be linked to the experience of the effect of water hardness on the lifetime of consumer appliances. Averting expenditures are found to decline with how long the respondent has owned the property, which could

[16] 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 2: Hardness of tap water – Household valuation function

	(I) Tobit model		(II) Tobit model with IV		
	β, γ	$\frac{\partial E(AVC\|AVC\geq 0,w,Z)}{\partial (w,Z)}$	First stage	β, γ	$\frac{\partial E(AVC\|AVC\geq 0,w,Z)}{\partial (w,Z)}$
RATING	74.88***	7.03***	–	129.55***	12.83***
	(14.25)	(1.22)		(23.38)	(2.52)
INCOME	1.21***	0.11***	0.001	0.91**	0.09*
	(0.43)	(0.04)	(0.001)	(0.44)	(0.05)
AGE	4.07***	0.38***	0.001	4.10***	0.41***
	(1.17)	(0.11)	(0.003)	(1.19)	(0.13)
FEMALE	-36.10	-3.40	0.07	-39.56	-3.94*
	(24.75)	(2.28)	(0.06)	(25.15)	(2.30)
HEALTHY	14.11	1.35	-0.18***	27.94	2.87
	(26.84)	(2.63)	(0.06)	(28.24)	(3.00)
HOSPITALIZED	-10.59	-0.96	0.04	-15.82	-1.50
	(37.90)	(3.33)	(0.09)	(38.38)	(3.43)
RESIDENCY	1.72	0.16	0.01*	1.30	0.13
	(2.40)	(0.22)	(0.01)	(2.44)	(0.24)
HOME OWNER	108.91*	10.68*	0.01	108.94*	11.24*
	(57.40)	(5.73)	(0.13)	(57.70)	(6.04)
RESIDENCY x OWNER	-6.50*	-0.61**	-0.01	-6.19*	-0.61*
	(3.35)	(0.30)	(0.01)	(3.35)	(0.32)
BILLS	-12.06	-1.13	0.02	33.04	3.27
	(79.62)	(7.48)	(0.18)	(78.53)	(7.72)
FAMILY SIZE	-2.05	-0.19	0.04	-2.83	-0.28
	(13.48)	(1.26)	(0.03)	(13.52)	(1.34)
INFANTS	-11.57	-1.04	-0.27	5.71	0.58
	(82.48)	(7.12)	(0.21)	(80.08)	(8.25)
MG $CaCO3/l$	–		0.008***	–	
			(0.0001)		
CONSTANT	-823.07***		1.73***	-1,018.52***	
	(132.46)		(0.18)	(157.27)	
σ	235.19***		–	230.60***	
	(25.67)			(24.58)	
ρ	–		–	-0.33***	
				(0.09)	
Log-(pseudo)likelihood	-1067.16		–	-2372.05	
(Pseudo) R^2	0.037		0.446	–	
N	1,029		1,029	1,029	
F-test/Wald (p-val)	0.00		0.00	0.00	
Wald stat. exo. (p-val)	–		–	0.00	
Partial 1^{st} stage F-stat.	–		802.21	–	

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. Observations are weighted for representativeness.

indicate an adaptation to the local level of water hardness. As should be expected for hardness-related expenditures the health indicators, family indicators and other controls are not found to have a statistically significant impact on averting expenditures.

Results for averting 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) is a simultaneous equation tobit in which the complaint rate is used to instrument the subjective aesthetic 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=0.123). 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. On the other hand, concerns about the complaint rate being a weak instrument also arise, as the partial F-statistic from the first stage regression (11.05) is near the cutoff values of Staiger and Stock (1997) and Stock and Yogo (2005). This is confirmed by evidence from the weak-instrument-robust 95% 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]. Therefore, 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% 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 Discussion of the exclusion restriction

Our exclusion restriction relies on the assumption that objectively measured water hardness and aesthetic quality are exogenous, in the sense that they affect averting expenditures only through perceived water harness and aesthetic quality. As discussed above, one challenge to the validity of the exclusion restriction is the possible endogeneity of objective water quality through preference-based sorting. This would imply that individuals with strong preferences for hardness or aesthetic quality choose to locate in WSZ with moderate water hardness and

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

	(I) Tobit model		(II) Tobit model with IV		
	β, γ	$\frac{\partial E(AVC\|AVC \geq 0, w, Z)}{\partial (w, Z)}$	First stage	β, γ	$\frac{\partial E(AVC\|AVC \geq 0, w, Z)}{\partial (w, Z)}$
RATING	-39.78***	-10.80***	–	-92.52***	-28.83***
	(3.24)	(0.80)		(34.34)	(10.62)
INCOME	0.39*	0.11*	0.004*	0.66**	0.20**
	(0.21)	(0.06)	(0.002)	(0.33)	(0.10)
AGE	0.43	0.12	0.002	0.56	0.18
	(0.52)	(0.14)	(0.005)	(0.61)	(0.19)
FEMALE	-38.31***	-10.47***	-0.139	-45.35***	-14.21***
	(10.73)	(2.95)	(0.106)	(12.99)	(4.01)
HEALTHY	5.33	1.46	0.380***	23.89	7.68
	(11.23)	(3.11)	(0.112)	(17.35)	(5.71)
HOSPITALIZED	25.37	7.57	-0.400**	-2.05	-0.64
	(16.36)	(5.33)	(0.173)	(26.78)	(8.25)
RESIDENCY	-0.61	-0.17	0.002	-0.42	-0.13
	(1.08)	(0.29)	(0.009)	(1.19)	(0.37)
HOME OWNER	29.73	8.17	0.231	52.24*	16.55*
	(23.43)	(6.54)	(0.232)	(30.53)	(9.82)
RESIDENCY x OWNER	-0.37	-0.10	-0.007	-1.38	-0.43
	(1.37)	(0.37)	(0.013)	(1.68)	(0.52)
BILLS	-7.34	-1.99	-0.033	-5.21	-1.62
	(33.69)	(9.15)	(0.341)	(38.64)	(12.04)
FAMILY SIZE	5.66	1.54	0.047	6.73	2.10
	(5.64)	(1.54)	(0.057)	(6.12)	(1.91)
INFANTS	-23.37	-5.69	-0.590	-68.95	-16.76
	(45.54)	(9.85)	(0.447)	(66.50)	(12.02)
FEMALE x INFANTS	82.46	32.01	0.776	166.36*	84.97
	(56.17)	(28.70)	(0.656)	(89.77)	(63.23)
COMPLAINTS	–		-0.629***	–	
			(0.189)		
CONSTANT	187.11***		7.318***	555.82**	
	(35.42)		(0.313)	(240.13)	
σ	126.50***		–	126.29***	
	(6.76)			(6.77)	
ρ	–		–	.74	
				(0.48)	
Log-(pseudo)likelihood	-2396.96		–	-4518.89	
(Pseudo) R^2	0.045		0.036	–	
N	1,074		1,074	1,074	
F-test/Wald (p-val)	0.00		0.00	0.00	
Wald stat. exo. (p-val)	–		–	0.123	
Partial 1^{st} stage F-stat.	–		11.05	–	

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. Observations are weighted for representativeness.

high aesthetic quality. Because these individuals would also undertake more averting measures, all other things equal, the exclusion restriction would be violated.

While it is of course not possible to test the validity of the exclusion restriction, it is important to discuss it's general validity in the context we consider. First, water hardness is determined by the geological characteristic of the water supply area, and water companies do not currently soften / harden water during treatment. This avoids any supply-side break-down of the exclusion restriction (e.g. Zivin et al., 2011), so that we can essentially think of water hardness as a randomly assigned feature of the water supply. Moreover, water hardness is not a salient feature of water quality, and it varies only over relatively large spatial areas, which would make sorting on this particular attribute particularly expensive for individuals. It is therefore safe to say that water hardness does not enter the location choice of individuals, ruling out preference-based sorting. As such the validity of our exclusion restriction is plausible for the case of water hardness.

For aesthetic features of the water supply, taste and odor in particular, the validity of the exclusion restriction is less straightforward. Indeed, while it is still highly unlikely that individuals' location decision depends on this particular amenity, the provision of these services depends on investments by local utility companies, and could thus be correlated with other socio-economic characteristics.[17] This issue comes up in the U.S. data on non-attainments of health-related water quality standards used by Zivin et al. (2011), as these service failures were found to be more likely to occur in areas with relatively low socio-economic outcomes.

One way to document a relationship between the provision of aesthetic water quality and socio-economic background of WSZ is to compare subpopulations who face different levels of service. In particular, we follow Zivin et al. (2011) and test whether subpopulations in areas below and above the median level of provision are observationally similar. Table 4 reports subsample means and t-tests for difference in means. Results suggest that there is no systematic variations across subsamples, except for the proportion of female that is marginally significant (p-value$<$0.1), although this hardly goes against the validity of the exclusion restriction. We also note that average income is higher in WSZ with complaints above the me-

[17] Note that this issue could potentially be mitigated by the use of historical data on objective quality. Unfortunately in the present empirical application we do not have access to such data.

dian, which goes against the evidence reported in Zivin et al. (2011), although the difference is not statistically significant.

The structure of water supply in England and Wales can at least partly account for this feature of our data. In particular, one key difference with the U.S. setting considered by Zivin et al. (2011) is that the water industry in England and Wales is composed of relatively large regional private monopolies that are regulated by the government.[18] The geographical coverage of each regional monopoly is relatively large, so that a household locating in a particular area would not have a choice between different suppliers, and information about WSZ boundaries is not observed (i.e. characteristics of treatment plants for different WSZ are not known). Moreover, regulation enforces strict compliance with European drinking water standards (99.96% compliance on standards related to human health, see DWI, 2013), which makes an association between service failures and local socio-economic conditions very unlikely.

In sum, the characteristics of water supply we consider are unlikely to affect households' location choices. While this is relatively uncontroversial for water hardness, service provision for aesthetic quality may be correlated with socio-economic characteristics, which would challenge the validity of the exclusion restriction. Based on differences on observables, however, we find little evidence against the exclusion restriction, a feature that can be related to the regulatory structure of water supply in England and Wales.

3.5 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 and waste water bills, health, household composition, and residency status), but remains nevertheless rather arbitrary. Moreover, the pseudo-R^2 statistics for tobit regressions suggest that the model only explains a fraction of variability in expenditures, which indicates that some key determinants of expenditure are not included. Potentially omitted determinants

[18] See Lanz and Provins (2015) for a detailed description of the regulation of water and waste water services in England and Wales.

Table 4: Sample composition by frequency of taste and odor complaints

	Complaints low	Complaints high	
	Mean	Mean	Difference
INCOME	31.69	34.17	-2.48
	(1.22)	(1.29)	(-1.40)
AGE	45.59	47.08	-1.49
	(0.87)	(0.89)	(-1.20)
FEMALE	0.53	0.46	0.07*
	(0.03)	(0.03)	(1.93)
HEALTHY	0.30	0.29	0.01
	(0.03)	(0.02)	(0.27)
HOSPITALIZED	0.09	0.08	0.01
	(0.13)	(0.14)	(0.26)
RESIDENCY	10.74	10.95	-0.21
	(0.46)	(0.44)	(0.34)
HOME OWNER	0.62	0.62	-0.001
	(0.02)	(0.02)	(-0.03)
BILLS	0.38	0.38	0.001
	(0.01)	(0.01)	(0.05)
FAMILY SIZE	2.37	2.38	-0.01
	(0.06)	(0.06)	(-0.13)
INFANTS	0.04	0.04	-0.003
	(0.01)	(0.01)	(0.25)

Notes: Complaints low and high subsamples refer to groups of respondents in WSZ respectively below and above the median level of complaints about taste and odor. Mean values are reported with standard deviations in parenthesis below. The column with "Difference" reports differences between subsamples means with t-statistics reported in parenthesis below. ***, **, *: statistically significant at 1, 5 and 10 percent respectively. Observations are weighted for representativeness.

are, in fact, an other reason why instrumenting perceived quality is important in order to estimate unbiased marginal effects. The second specification provides evidence on the potential bias imposed by the assumption of joint normality of outcome variables associated with the simultaneous-equation tobit model.

To assess the impact of covariate selection, we consider the 'minimal' specification required to obtain marginal WTP estimates. Specifically we estimate a model in which the only covariate included is the rating variable. Table 5 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. The simultaneous-equation tobit model still suggests

Table 5: Hardness of tap water – Minimal valuation function

	(I) Tobit model		(II) Tobit model with IV		
	β, γ	$\frac{\partial E(AVC \mid AVC \geq 0, w, Z)}{\partial (w, Z)}$	First stage	β, γ	$\frac{\partial E(AVC \mid AVC \geq 0, w, Z)}{\partial (w, Z)}$
RATING	79.87***	8.15***	–	132.24***	14.00***
	(15.72)	(1.47)		(24.69)	(2.64)
MG $CaCO3/l$	–		0.008***	–	–
			(0.0001)		
CONSTANT	-584.99***		2.01***	-766.40***	
	(86.62)		(0.06)	(118.35)	
σ	246.77***		–	242.68***	
	(28.36)			(27.59)	
ρ	–		–	-.30***	
				(0.08)	
Log-(pseudo)likelihood	-1084.3		–	-2399.44	
(Pseudo) R^2	0.021		0.439	–	
N	1,029		1,029	1,029	
F-test/Wald (p-val)	0.00		0.00	0.00	
Wald stat. exo. (p-val)	–		–	0.00	
Partial 1^{st} stage F-stat.	–		850.96	–	

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. Observations are weighted for representativeness.

that perceived quality is endogenous, and the coefficient of the auxiliary regression and results for the marginal WTP estimates barely change (the latter increases from 12.83 with the full list of covariates to 14.00 in the model with no covariates).

Table 6 provides similar evidence for aesthetic quality, as marginal WTP estimates from the simultaneous-equation tobit model increases from 28.83 with covariates to 33.79 without. One important finding, however, is that removing covariates from the aesthetic quality specification 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 20, while at the same time the coefficient from the first stage equation remain close to those in Table 3. Bounds from the robust 95% confidence interval [-180.3,-30.7] are also much closer to those of the traditio-

Table 6: Aesthetic quality of tap water – Minimal valuation function

	(I) Tobit model		(II) Tobit model with IV		
	β, γ	$\frac{\partial E(AVC \mid AVC \geq 0, w, Z)}{\partial (w, Z)}$	First stage	β, γ	$\frac{\partial E(AVC \mid AVC \geq 0, w, Z)}{\partial (w, Z)}$
RATING	-38.84***	-10.72***	–	-103.17***	-33.79***
	(3.32)	(0.83)		(38.98)	(12.67)
COMPLAINTS	–		-0.65***	–	–
			(0.15)		
CONSTANT	212.15***		7.61***	692.07**	
	(21.12)		(0.08)	(289.93)	
σ	130.47***		–	130.14***	
	(7.26)			(7.26)	
ρ	–		–	0.88*	
				(0.53)	
Log-(pseudo)likelihood	-2415.23		–	-4554.74	
(Pseudo) R^2	0.038		0.01	–	
N	1,074		1,074	1,074	
F-test/Wald (p-val)	0.00		0.00	0.00	
Wald stat. exo. (p-val)	–		–	0.097	
Partial 1^{st} stage F-stat.	–		20.07	–	

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. Observations are weighted for representativeness.

nal Wald confidence interval [-156.9, -26.7]. 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 toward zero.

A second source of concern with the simultaneous-equation tobit model is related to the assumption of joint normality of errors. 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 averting expenditure data. Results reported in Table 7 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 13.02, and 12.83 from the

Table 7: Linear 2SLS estimation – Household valuation function

	(I) Water hardness	(II) Aesthetic quality
RATING	13.02***	-25.65**
	(2.70)	(11.17)
CONSTANT	-66.09***	195.74**
	(15.91)	(78.34)
R^2	0.083	0.142
N	1029	1074
F-test (p-val)	0.00	0.00
K-P F-stat.	799.77	7.71

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. Observations are weighted for representativeness.

simultaneous-equation tobit model (see Table 2). For aesthetic quality, marginal WTP from the 2SLS model is -25.65, which is slightly lower than that derived from the simultaneous-equation tobit model, -28.83 (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 validity of marginal WTP estimates derived from the simultaneous-equation tobit model. This is especially the 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.

3.6 Marginal willingness to pay and objective quality

We now finish the presentation of our empirical illustration with a more detailed discussion of marginal WTP estimates. First, Table 8 provides further evidence about marginal WTP for water softening and aesthetic quality based on simultaneous-equation tobit specification reported in Table 2 and 3. Specifically, marginal effects are highly non-linear functions of the estimated parameters and thus vary with the point at which they are evaluated. Here we consider how marginal WTP for improvements in service rating varies across different income levels and baseline service ratings.

Marginal WTP for water softening is found to be substantially higher (around GBP 20 per household per year) when water is rated to be relatively hard (i.e. when the hardness rating variable is high) as compared to when it is rated to be soft (around GBP 3 per household per year), which is as expected. Similarly, marginal WTP for improvements to taste and odor is around GBP 90 per household per year when the rating variable is near its minimum value, and declines to around GBP 20 per household per year when rating is near it's maximum value.

Second, we use results from the auxiliary regression relating objective quality and service rating to evaluate marginal WTP for changes in objective quality. Typically, objective quality is the outcome targeted by investments in infrastructures, whereas perceived quality would be affected only indirectly. Here we illustrate how our approach can be used to derive marginal WTP evidence associated with improvements in objective quality both at the average of the sample and for an improvement from the minimum level of objective quality reported in our sample to the corresponding average value. Note that the latter estimates are derived by taking the difference in average expenditures predicted at two different values of the objective quality variable (through its effect on rating variables), while other variables are kept at their sample mean. These figures thus account for the fact that marginal WTP estimate vary with the point at which they are evaluated (as shown in Table 8). The associated robust standard errors are evaluated using the delta method.

For water hardness, a unit change in objective quality (1 mg CaCO3 per liter) amounts to a change in the rating variable of around 0.008. In turn, this translates into a marginal WTP of GBP 0.106 per household per year (p-value<0.01; 95% confidence interval: 0.066 – 0.147). Furthermore, maximum water hardness measured in our sample is 387.75 mg CaCO3 per liter, and the difference with the sample average (168.76) is 218.99. Using the estimated relationship between objective and subjective variables, this corresponds to an decrease in the hardness rating variable of around 1.8. From the second stage regression, this amounts to a decrease in average averting expenditure of GBP 49.39 per household per year (p-value<0.01; 95% confidence interval: 25.28 – 73.51). This is a significant amount, as it represents about

Table 8: Marginal willingness to pay for water softening and aesthetic quality improvements

		mWTP (GBP/household/year) $\frac{\partial E(AVC \mid AVC \geq 0, w, Z)}{\partial w}$		
Yearly income ('000GBP)		25	40	65
Hardness rating	2	2.60*	2.97**	3.70**
		[-0.01 - 5.20]	[0.11 - 5.84]	[0.25 - 7.15]
	3	8.33***	9.29***	11.07***
		[4.17 - 12.49]	[4.99 - 13.59]	[6.10 - 16.04]
	4	20.98***	22.83***	26.12***
		[13.03 - 28.94]	[14.76 - 30.89]	[17.01 - 35.22]
Aesthetic rating	3	90.95**	91.18**	91.49**
		[17.31 - 164.58]	[18.19 - 164.16]	[19.49 - 163.50]
	5	76.43*	77.88*	80.09*
		[-3.69 - 156.55]	[-2.70 - 158.46]	[-0.90 - 161.09]
	8	18.75***	20.43***	23.42***
		[11.37 - 26.13]	[11.16 - 29.70]	[10.25 - 36.59]

Notes: 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; 95% confidence intervals reported in brackets. Observations are weighted for representativeness.

28% of average water bills in England and Wales,[19] although one should bear in mind that this marginal WTP estimate only relates to households in WSZ with particularly hard water.

For aesthetic water quality, our objective measure captures the number of complaints per thousand of customers in each WSZ, and a unit reduction in our complaints measure is associated with an increase of the rating variable of around 0.63, which translates into a marginal WTP estimate of GBP 18.13 per household per year (p-value<0.01; 95% confidence interval: 5.04 – 31.23). Further considering a decrease from the maximum rate of complaints (1.71) to the average value in our sample (0.43), which corresponds to an increase of the rating variable for taste and odor of about 0.8, we find that the marginal WTP estimate is GBP 28.02 per household per year (p-value=0.036; 95% confidence interval: 1.76 – 52.28). In other words, WTP corresponding to a change from maximum complaint rate to the sample average is equivalent to around 15% of yearly household water bills.

[19] Water bills in England and Wales, excluding waste water services, amount to GBP 180 per household per year; see Ofwat (2013).

4 Concluding comments

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 averting 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, we 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 auxiliary (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, we have used data on averting expenditures for water hardness and aesthetic quality of tap water. First, we 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 averting expenditure are relatively minor in terms of overall household budget, it is still significant 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 products improving perceived water quality, 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 relationship between objective and perceived quality (the 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, we had to rely on regional complaint rates as a measure of objective quality, which raises 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 generates interesting and robust results. 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 in 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.

——— (2016) "The impact of perceptions in averting-decision models: An application of the special regressor method to drinking water choices," *American Journal of Agricultural Economics*, 98 (1), pp. 297–313.

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.

Chay, K. Y. and M. Greenstone (2005) "Does air quality matter? Evidence from the housing market," *Journal of Political Economy*, 113, 2, pp. 376–424.

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 (2015) "Using discrete choice experiments to regulate the provision of water services: Do status quo choices reflect preferences?" *Journal of Regulatory Economics*, 47 (3), pp. 300 – 324.

——— (2016) "The demand for tap water quality: Survey evidence on water hardness and aesthetic quality," *Water Resources and Economics*, `http://dx.doi.org/10.1016/j.wre.2016.10.001`.

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/pn-0112-water-and-sewerage-bills-to-rise-in-2012/` (accessed March 2016).

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.

Poe, G. and R. Bishop (1999) "Valuing the incremental benefits of groundwater protection when exposure levels are known," *Environmental and Resource Economics*, 13 (3), pp. 347–373.

Poe, G., H. van Es, T. Van den Berg, and R. Bishop (1998) "Do households update their health risk and exposure perceptions with water testing?" *Journal of Soil and Water Conservation*, 53 (4), pp. 320–325.

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.

Shogren, J. and T. Crocker (1991) "Risk, self-protection, and ex ante economic value," *Journal of environmental economics and management*, 20, pp. 1 – 15.

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.

Tiebout, C. M. (1956) "A pure theory of local expenditures," *Journal of Political Economy*, 64 (5), pp. 416 – 424.

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.