CrossMark

Responses to Changes in Domestic Water Tariff Structures: A Latent Class Analysis on Household-Level Data from Granada, Spain

María Pérez-Urdiales · **María A. García-Valiñas** ·
Roberto Martínez-Espiñeira

Abstract A problem that affects the estimation of water demand functions is the presence of unobserved individual heterogeneity, which means that a common demand function is unlikely to represent the behavior of all users. We implement Latent Class Models to estimate water demand functions for four groups of users who are classified according to their unobservable preferences. This more flexible approach makes it possible to distinguish four different response patterns to changes in the drivers of water use, including different price elasticities. These results should be of particular interest to regulators who would like to tailor water demand management policy to heterogeneous users. Our analysis exploits household-level panel data on residential water demand and consumers' characteristics obtained by combining information from a survey of 1,465 domestic users in the city of Granada and bimonthly price and consumption data supplied by this city's water supplier from the period 2009–2011.

1 Introduction

Demand for water is expected to increase during the next few years as the world population is also predicted to grow from 6.9 billion in 2009 to 8.3 billion in 2030 and 9.1 billion in

M. Pérez-Urdiales (✉)
Facultad de Economía y Empresa, University of Oviedo, Av. del Cristo s/n, 33006 Oviedo, Spain
e-mail: perezumaria@uniovi.es

M. Pérez-Urdiales · M. A. García-Valiñas
Oviedo Efficiency Group, University of Oviedo, Oviedo, Spain

R. Martínez-Espiñeira
Department of Economics, Memorial University of Newfoundland, St. John's, Canada

🌐 Springer

2050 (UNDESA 2009). This population growth and the increasing urbanization trend will lead to a higher water demand and, at the same time, compromise the ability of ecosystems to provide conventional and cleaner supplies (The World Bank 2012). Although there are different types of strategies to deal with imbalances between water supply and demand, the use of demand-side policies has emerged as a preferred option during the last decades. Among these, pricing policies have become a particularly attractive option, since they may result in lower levels of efficiency losses than other rationing alternatives (Roibás et al. 2007). In this sense, accurate estimates of price elasticity of water demand are crucial for policy decision-making, since they make it possible for water policy designers to understand how strongly water consumption will react to changes in price. However, due to the characteristics of the good, pricing policies are often constrained by regulation, due to the characteristics of the good (Olmstead and Stavins 2007), so water suppliers also use non-price conservation programs to induce water reductions.

In the past, the neoclassical approach assumed that "tastes neither change capriciously nor differ importantly between people" and it was limited "to search[ing] for differences in prices and incomes to explain any differences or changes in behavior" (Stigler and Becker 1977, p. 76). However, there may be great heterogeneity in water consumption, even amongst individuals who are similar in observable characteristics. Therefore, when analyzing the estimation of residential water demand using microdata, addressing unobserved heterogeneity is a critical issue, since the demand functions are influenced by unobservable heterogeneous preferences.

During the last few years, the problem of unobserved heterogeneity has received special attention, resulting in two main ways of addressing it. One approach confines unobserved heterogeneity in an individual-specific effect, as in linear panel data models such as fixed-effects and some random-effects models, while assuming that the marginal response to the demand determinants is the same across individuals. A more flexible approach, based, for example, on the use of random coefficients models, assumes, instead, that the regression parameters vary randomly across individuals according to some distribution and identifies the mean and the standard deviation for these parameters (Cameron and Trivedi 2005, pp. 9–10). Another common strategy in the water demand literature has been to group individuals a priori according to observable characteristics that are assumed to be proxies for unobserved preferences and tastes.

The approach adopted here differs from the previous literature in that we use a Latent Class Model (LCM) to control for heterogeneity in preferences. It allows us to identify a finite number of consumer "classes" and hence different water demand functions. This methodology consists of estimating the model in two simultaneous steps. One involves estimating the main function, in this case water demand, and the other step estimates the probability that each consumer to belong to a given class. By sorting individuals based on the similarity of their unobserved component, the LCM accommodates unobserved heterogeneity, while tractability and theoretical consistency are preserved in terms of the so-called Ockham's razor. That is, water use would be perfectly explained with a model in which each consumer had a different water demand function. However, that model would be practically intractable and useless for predicting water consumption. Instead, the LCM groups consumers into the minimum number of classes that is consistent with common preferences.

The LCM has several advantages over the techniques discussed above. Compared to the first approach (using linear panel data models, such as fixed-effects and random-effects) it also accounts for slope heterogeneity across different groups of consumers, instead of confining unobserved heterogeneity to an individual-specific effect and constraining all consumers to have the same marginal effects. As stated above, the second approach (the random coeffi-

cient model) assumes that the coefficients are different for each consumer. On the other hand, the LCM identifies consumer profiles that may be more easily managed when it comes to effecting water conservation policies. The LCM does not require making an *ad hoc* selection to the membership, which could be highly sensitive to arbitrariness, since it segments consumers endogenously into different groups. Moreover, the LCM identifies classes and allows to flexible modelling of the probability of belonging to a certain group (within which unobservable preferences are similar) as a function of a set of, in principle, observable covariates. Therefore, it provides information about the size of each group and a description of the type of consumer belonging to them. This information can be very useful for the design of water management policies, as long as information about the factors that affect class membership can be obtained at a reasonable cost.[1]

Our application exploits a panel dataset from Granada (Spain) that contains information on water consumption and prices for the period 2009–2011, as well as on socioeconomic variables and self-reported water conservation habits from a household survey carried out in 2011, which can be useful to control for individual heterogeneity. This data set is of particular interest for two reasons. First, Spain is the most semi-arid country in the European Union (Lopez-Gunn et al. 2012) and the South of Spain, where the city of Granada is located, is regularly affected by droughts and other water availability issues. Thus, it is important to understand residential water demand in order to improve water management. Second, there was a change in the price structure in the city of Granada in 2011, which makes it possible for us to consider not only changes in price levels but also changes in the size and number of price blocks when analyzing consumer responses to the water tariff. Our findings provide potentially useful information for regulators by identifying four different residential water consumer profiles. We also derive some rather informative conclusions from the analysis of the change in the price structure effected in 2011. Additionally, a sensitivity analysis is also carried out to compare the results obtained using the LCM with those obtained using an alternative grouping technique. This analysis illustrates the superiority of LCMs for identifying homogeneous groups of consumers.

The paper has the following structure. In Sect. 2, we discuss different methods that previous literature has applied to deal with heterogeneity issues. Section 3 presents the econometric model. Section 4 describes the tariff structure in the city of Granada, paying special attention to the change in the structure in 2011. Section 5 describes the data. Estimates from the LCM and sensitivity analysis are presented in Sect. 6. Finally, Sect. 7 concludes summarizing the main results.

2 Background

Understanding residential water demand is essential to the effective management of water resources. Consequently, the literature on residential water demand is vast, as revealed by the many studies that have surveyed the estimation of water demand. For example, Arbués et al. (2003) focus on different modeling approaches and data sets; Dalhuisen et al. (2003) include a meta-analysis of price and income elasticities; Worthington and Hoffman (2008) provide

---

[1] On the other hand, the advantages of using LCMs could be not worthwhile if, after the process of estimating the model, water managers uncovered the determinants of class membership to be unobservable (that is, practically unobservable, rather than just part of the error component of the water demand functions) or that obtaining data about those determinants imposed a large informational burden. We thank an anonymous referee for pointing this out.

a survey of model specification and results; and Nauges and Whittington (2010) review the literature analyzing household residential demand in developing countries.

As mentioned in Sect. 1, it is important to account for heterogeneity, particularly when analyzing the effect of a change in the price structure. Differences in terms of price elasticities may be due to the underlying heterogeneity among regions and even households. Thus, an increasing number of studies aim to control for the presence of unobserved individual heterogeneity in residential water demand. However, the common methods to address heterogeneity seem to perform relatively poorly under certain circumstances. We will discuss the problems associated to each technique, providing additional arguments to support the use of Latent Class Analysis.

A frequently adopted approach is to control for unobserved household heterogeneity through the inclusion of household fixed effects. For example, Pint (1999) uses a fixed-effects model and an ordinary least squares (OLS) model to estimate household responses to water price structure changes in California, finding the fixed-effects model to be preferable to the OLS model. However, none of these estimations considered instrumental variable (IV) specifications, resulting in upward-sloping water demand at high prices. Worthington et al. (2009) analyze residential water demand in several councils in Queensland by estimating *common-effects* (whereby they assumed that water consumption was homogeneous across local councils), fixed-effects, and random-effects models. Their results show that the fixed-effects model outperforms the others for that particular case. Coleman (2009) develops dynamic models of water demand in Salt Lake City estimated using fixed-effects models and compares them with static models obtained using pooled, fixed-effects and random-effects models. Polebitski and Palmer (2010) estimate pooled, fixed-effects and random-effects models to analyze single-family residential water demand for over 100 census tracts for the period 1991–2005, and the Hausman test indicates that the fixed-effects model is preferred over the random-effects model. Nataraj and Hanemann (2011) include household and year fixed-effects into a regression discontinuity model to account for heterogeneity across the treatment and control households in a natural experiment to determine whether consumers react to an increase in marginal price.

As previously discussed, another way to handle heterogeneity is through the estimation of Random-Coefficient Models (RCM). This methodology has not been widely used in the water demand literature, llikely due to its difficult interpretation as a tool to identify groups of individuals with relatively similar responses to changes in dependent variables. This is because the RCM assumes a continuous distribution of random individual-specific regression parameters and only identifies the mean and the standard deviation of each of these distributions. This limits its usefulness in a case like ours. As far as we are aware, Miyawaki et al. (2010) is the only study that applies this methodology in this field. They conduct an analysis of Japanese residential water demand estimating a random parameters model and an autoregressive of order one error component model, obtaining similar results from both estimations.

An alternative approach consists of including dummy variables to indicate socioeconomic and demographic characteristics that can capture differences in individual's preferences, in the demand function. Renwick and Green (2000) incorporated irrigation dummy variables into the demand equation to account for differences in outdoor water use. Krause (2003) investigated consumer heterogeneity in water demand using a set of experiments and a survey. First, they included group dummy variables interacted with the parameters in the demand function and then computed disaggregated demand functions for three consumer types considered in the experiment and surveys: students participants, workforce participants and

retired participants. Therefore, the ability of this technique to control for heterogeneity using this technique is clearly limited.

Some studies identify different groups of consumers according to observable characteristics that may be related to the consumers' unobserved preferences. Renwick and Archibald (1998) analyze the effect of demand side policies by clustering groups of consumers in terms of income. Ruijs et al. (2008) estimate a linear demand function in the Metropolitan Region of Sao Paulo for the period 1997–2002 and evaluate welfare and distribution effects for five income groups. Mansur and Olmstead (2012) divide the sample into four sub-groups based on income and lot size in order to compare different price elasticities for indoor and outdoor water demand. However, these techniques make an ad hoc selection to the membership, which is highly sensitive to arbitrariness.

LCMs have attracted increased attention lately since, as we will describe in Sect. 3, this technique presents significant advantages. Among these advantages, we exploit the fact that it makes it possible to generate homogeneous groups of consumers without setting any *a priori* criteria. A number of studies use this methodology to analyze demand in other economic fields, such as health economics (Deb and Trivedi 2002; d' Uva 2006; Ayyagari et al. 2013; Hyppolite and Trivedi 2012), cultural economics (Boter et al. 2005; Fernandez-Blanco et al. 2009; Grisolía and Willis 2012) or transport (Hensher and Greene 2003; Shen et al. 2006; Shen 2010; Hess et al. 2011; Greene and Hensher 2013).

There are several applications of the LCM in environmental economics. For example, Scarpa et al. (2005) compare the use of the mixed logit random parameter model with the use of Latent Class Analysis to model the choice of water utility by the consumer. Patunru et al. (2007) implement this methodology to investigate the willingness-to-pay for the clean-up of hazardous waste by homeowners in Waukegan, Illinois. Scarpa et al. (2007) study different groups in the demand for hiking in the eastern Italian Alps, arguing that it is fundamental to assess heterogeneity when analyzing expected consumers surplus, predicted visitation, and response to access fees. Campbell et al. (2011) identify heterogeneous groups of respondents that were asked about the willingness-to-pay for improvements in four rural landscapes in the Republic of Ireland. However, to our knowledge, there have been no applications as yet to residential water demand functions.

Another typical concern identified in the residential water demand literature is about price endogeneity, especially in the presence of nonlinear prices. As detailed by Olmstead (2009b), there are two types of estimation approaches that have been used in the literature to control for this problem: reduced-form approaches, such as IV, and structural approaches, such as discrete/continuous choice models (DCC). The IV approach is often undertaken in water demand analysis along with fully parametric or semiparametric methodologies as two-stage least squares (2SLS) or Generalized Method of Moments (GMM).[2]

In DCC models, a consumer's utility maximization problem is solved in two steps. First, the consumer selects the block given the price of each block and then decides the level of consumption that maximizes her utility. These models have been used by relatively few papers in this literature. Hewitt and Hanemann (1995) develop a DCC model of residential water demand using household level data from Denton (Texas) for the period 1981–1985, obtaining price elasticities in the range of −1.57 to −1.63, which are much higher than those obtained in the literature based on using IV techniques. Apart from the models described above, Pint (1999) applies DCC models obtaining relatively low price elasticities. Olmstead (2009b) compares IV and DCC estimates of water demand under increasing-block pricing

---

[2] See for instance, Agthe et al. (1986) in IV, Nieswiadomy and Molina (1988, 1989) in 2SLS or García-Valiñas (2005) for GMM.
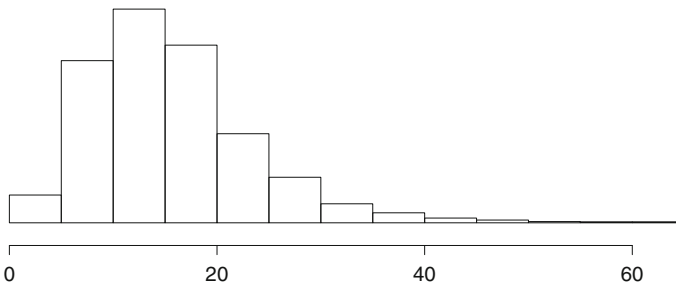
Fig. 1 Distribution of residential water demand in Granada 2009–2011

using a Monte Carlo experiment finding that both models exhibit significant bias in the simulations. Strong and Smith (2010) critique DCC models, as Bockstael and McConnell (1983) stated that the Marshallian "prices as parameters" demand function does not exist with a nonlinear budget constraint and, therefore, applied welfare analysis is problematic in this case. Moreover, this model is based upon marginal prices, which assumes that consumers are aware of the price structure.[3] There is no general theory recommending how to control for endogeneity in LCMs. Nevertheless, we use a two-stage control function approach (explained in the "Appendix 1") because it performs better on nonlinear models.

3 Methodology

From a methodological viewpoint, LCMs are proposed to identify different groups of consumers. This methodology may perform well to estimate residential water demand for two main reasons. First, water demand functions are related to utility functions, which are based on consumers' unobservable preferences and tastes that may differ across consumers. Therefore, LCMs allow us to identify groups of consumers who have similar unobservable preferences about how to change their water use in response to changes in a certain set of observable explanatory variables, since it sorts individuals based on the similarity of their conditional distributions. Second, from a statistical point of view, Fig. 1 shows that the distribution of residential water consumption is asymmetric in our sample. Therefore, this distribution is better approximated by a mixture of several normal distributions rather than a single normal (and symmetric) distribution.

In LCMs, we assume that the sample of individuals is drawn from a population that is a finite mixture of C distinct subpopulations (Cameron and Trivedi 2005). The density of the dependent variable (residential water consumption) y, for observation i conditionally on some parameters (β, π) and on some explanatory variables x can be written as:

$$f(y_i|x_i; \beta, \pi) = \sum_{j=1}^{C} \pi_j f_j(y_i|x_i\beta_j), \quad i = 1, \ldots, n \tag{1}$$

where π_j is the probability of choice j of individual i ($\sum_{j=1}^{C} \pi_j = 1$ and $\pi_j \geq 0$, $j = 1, \ldots, C$).

---

[3] However, this assumption does not hold for our sample, since only 34.62 % of the households know the price schedule they face.

If any potential sources of heterogeneity are observed, the probability that consumer i belongs to class j can be parameterized as a function of covariates assuming that the latent variable follows a multinomial probability that yields a multinomial logit model:

$$\pi_j = \frac{exp(\gamma_j' z_i)}{\sum_{j=1}^{J} exp(\gamma_j' z_i)}, \quad j = 1, \ldots, J \tag{2}$$

where γ_i is a vector of parameters to be estimated and z_i is a vector of observable characteristics and self-reported valuations that may be considered proxies for the underlying utility preferences (Fernandez-Blanco et al. 2009).

Therefore, if we consider a normal mixture, the log-likelihood is defined as the sum of C log-likelihood normal distributions weighted by the probabilities of class membership:

$$\mathfrak{L}(\beta, \gamma) = \sum_{j=1}^{C} P_{ij} \frac{1}{\sqrt{2\pi\sigma_j^2}} exp\left(-\frac{1}{2\sigma_j^2}\left(y_i - x_i \beta_j^2\right)\right) \tag{3}$$

One of the key issues in the application of the LCM is how to correctly determine the number of classes. Although LCMs with additional numbers of classes are considered nested models, it is not possible to identify the correct model using a likelihood ratio test (LRT), because regularity conditions are not met (Nylund et al. 2007). The usual way to proceed is to estimate models with increasing numbers of classes in a stepwise fashion and compare the results using likelihood-based information criteria, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC). However, since these criteria do not share the same properties, they may yield contradictory verdicts. Nylund et al. (2007) analyzed the performance of these information criteria using Monte Carlo simulations and found that the Bayesian Information Criterion outperforms the others in correctly identifying the optimal number of classes.

Once the model is estimated, we use the parameter estimates to compute the posterior probabilities of belonging to each latent class:

$$Pr[y_i \epsilon c | x_i; y_i; \theta] = \frac{\pi_c f_c(y_i | x_i; \theta_c)}{\sum_{j=1}^{C} \pi_j f_j(y_i | x_i; \theta_j)} \quad c = 1, \ldots, C \tag{4}$$

4 Residential Water Tariffs in Granada

The water pricing structure in the city of Granada is based on increasing block prices (IBP). In this case, the tariff[4] also includes a fixed water service fee that must be paid regardless of the level of use and a set of increasing block prices. As shown in Table 2, the price structure in Granada remained unchanged between 2009 and 2010 but in 2011 the size of the price blocks was altered.

The fixed component of the tariff includes a water supply fee, a sewage collection fee, and a treatment fee and, in 2009 and 2010, a drought surcharge. Additionally, in 2011 a water tax collected on behalf of the Andalusian regional government was incorporated into the tariff.

The evolution of the prices in each block is shown in Table 1 (in real terms calculated using the province-level Consumer Price Index with base 2011).

---

[4] The tariff also includes discounts to those who are unemployed, retired, or have a certain minimum number of dependants.

Table 1 Evolution of prices 2009–2011

Year	P. Block 1(€)	P. Block 2(€)	P. Block 3(€)	P. Block 4(€)	P. Block 5(€)
2009	0.9798	1.9130	1.9310	2.4451	2.7356
2010	1.0318	1.9365	1.9545	2.5411	2.9137
2011	0.9731	1.3536	2.3534	3.4347	–

Table 2 Evolution of the size of pricing blocks

Blocks	2009–2010 (m^3)	2011 (m^3)
Block 1	0–8	0–2
Block 2	8–10	2–10
Block 3	10–16	10–18
Block 4	16–30	>18
Block 5	>30	–

Table 3 Evolution of the average total bill and the average quantity of water consumed

	2009	2010	2011
Water consumed[a] (m^3)	15.4939	16.0069	15.2579
Total bill (€)	44.3969	45.0625	49.1680

[a] Water consumed per household every 2 months

The tariffs in Granada were reviewed annually. Block prices were adjusted upwards from 2009 to 2010 but, as mentioned above, the price structure remained unchanged. However, in 2011 the rate schedule was also changed. [5] The rate schedule is described in Table 2.

As water becomes increasingly scarce in the South of Spain, water supply managers are using price as a water conservation tool. As stated above, Granada experienced a change in the price structure that resulted in a decrease in average water consumption but also an increase in the average total bill (Table 3).

5 Data

Our dataset is an unbalanced panel of bimonthly observations corresponding to 1,465 households in the city of Granada covering the period 2009–2011. The data come from two sources. The first source of information consists of water consumption and water tariffs data on a random and representative sample of urban households in the city of Granada, provided by EMASAGRA, the company in charge of water supply and sewage collection in Granada. The second one is a 2011 survey of these households, who were questioned about socioeco-

---

[5] To the best of our knowledge, there is only one previous work which deals with a change in the price structure similar to the one exploited in this paper. Martínez-Espiñeira and Nauges (2004) study residential water demand in Seville (Spain) for the period 1991–1999, having a slight change in the block size from 1996. Water demand is modeled using Stone–Geary utility function that allows identifying a threshold of water that is insensitive to price, however the change in the block is not directly analyzed.

nomic characteristics (occupation, household size), housing characteristics (size, equipment), attitudes towards the environment, and conservation habits.

Data on water consumption and water tariffs was merged with survey data. Since the survey was carried out in 2011, we only have information related to socioeconomic characteristics from that year. However, since the variables considered in the analysis are usually time invariant in the short and mid-run, we consider them applicable to the period 2009–2011.

Moreover, the average price is the only time-varying explanatory variable in our specification. As explained in Sect. 4, the pricing structure is based on IBP and, therefore, we must consider the price endogeneity generated by the simultaneous determination of the price level and the level of consumption that determines the price block. When addressing this issue, we face the problem that both water consumption and average price change within a given year in our dataset but we cannot observe a set of exogenous instruments that also vary within each year. Therefore, in order to address the endogeneity problem, it was necessary to aggregate water consumption by year, which made it possible to use the set of marginal prices per block, which change yearly, as instruments. Furthermore, after this transformation, we also excluded from the sample those individuals who were not observed for the entire year 2011, because of the possible bias introduced by seasonality in their water consumption.

Therefore, after the data aggregation, the dependent variable in our specification is the average bimonthly household water consumption per year, in cubic meters, which was calculated by dividing total consumption per year by the number of 2-month billing periods. Regarding the price variable, there are two main issues related to price when analyzing water demand facing a nonlinear pricing scheme. First, one must face the choice between marginal and average price. In this particular case, consumers indicated that they were not well informed about the pricing scheme. Therefore, households may be more sensitive to changes in average price (*AvP*) than in marginal price. The second issue, as commented above, relates to the price endogeneity generated by the simultaneous determination of price and the block of consumption. In the absence of a general theory about how to handle endogenous explanatory variables in LCMs, we used a two-stage control function estimation technique (CF) (Blundell and Powell 2003; Imbens and Wooldridge 2007; Howard and Roe 2013) over two-stage least squares (2SLS), because it is more appropriate for nonlinear models and, although our model is linear in parameters (since we are estimating mixtures of normal distributions), the nonlinearities arise when estimating the posterior probabilities at each maximization stage. The 2SLS approach would fail, because it implies approximating the endogenous variable with a linear transformation thereof and then using the estimated coefficients in the second stage would be used to compute the posterior probabilities. Therefore the nonlinear step in the computations of the posterior probabilities is not invariant to the use of 2SLS. A more detailed discussion of this methodology and the instruments selected can be found in the "Appendix 1" and the results of the estimation.

In order to allow for the possibility that the price elasticity differs between 2009–2010 and 2011, that is, between the period before and after the change in the price structure, we include an interaction (labelled *AvP2011*) between binary indicator for 2011 (*Year2011*) and the average price (*AvP*). Following Gaudin (2006), we incorporated also an indicator about how well-informed users are about their water tariff in the demand equation through its effect on price elasticity. To do so, we interacted a binary variable indicating awareness of the price structure with the average price (resulting in variable *Priceinfo*).

Household income was recorded as an ordered categorical variable, with households belonging to one of the following intervals (in Euros/month): [0–1,100]; [1,101–1,800]; [1,801–2,700]; [2,701–3,500]; [3,501–+∞]. It would not be appropriate to use the interval categories as if they were values of a continuous variable. Usually, one would construct a set of

five binary indicators of income level and introduce four in the model. However, because we did not seem to have enough sample variability to estimate all four corresponding parameters, we simplified our original income variable into a binary indicator (*Highincome*) of relatively higher income. In particular, we created a binary variable that identifies the richer households (those falling in the two highest income categories).

Additionally and based on previous literature, there were other variables included in the demand function. Household size (*Members*) was included, following most previous studies of residential water demand. Water conservation habits (*Habits*) were included using an aggregate index based on different daily behaviors,[6] as well as several variables representative of housing equipment are also considered, such as the number of electrical appliances (*Electappl*) and the number of efficient water-using electrical appliances (*Electeff*). Finally, an indicator of home ownership was included, since homeowners are expected to have more incentives than tenants to make investments in water-saving devices in the property, as shown by Grafton et al. (2011).

As stated in Sect. 3, individuals are sorted by the LCM into groups based on the similarity of their conditional distributions. However, the probability of belonging to a certain group can be further modeled as a function of covariates that can be considered proxies for the unobserved preferences related to water demand. That is, these variables allow us to identify household characteristics for the different water user profiles. In this sense, these variables belong in the class-selection function and not, or not so prominently, in the water demand functions. This is to say, they determine how the households' quantity demanded reacts to demand drivers, in particular how prices affect the quantity demanded of water, but not so much the quantity demanded as such.

At this stage, water suppliers would benefit from the knowledge of the main factors determining class membership, in particular the membership to classes with some specific characteristic (for example, a particularly (in)elastic response to price changes or a particularly sensitive response to moral suasion campaigns), if combined with the availability of individual data about those variables.

In practice, these variables could end up being easily observable factors (e.g. household type, presence of children, living in an apartment or not, availability of an individual meter, being a year-round versus a seasonal dweller, education levels, renter status, etc.) but they might instead be variables whose values would be costly for the water supplier to gather. The practical advantages afforded by the application of the LCM to estimate water demand would be dependent on these informational requirements, apart from having the technical sophistication and computational resources to periodically estimate the LCM itself.[7] In some cases, a more *ad hoc* approach to identifying consumer groups might thus end up being more efficient, particularly in those cases in which the differences among classes in terms of the most relevant estimated parameters (such as price-elasticities) are minor.

Following Russell and Fielding (2010), we use Stern (2000) as a guide when categorizing the determinants of different water demand behaviors into four types of causal variables.

Attitudinal factors are one of the causes of behaviors. According to the value-belief-norm theory (VBN), "the general predisposition to act with proenvironmental intent can influence all behaviors an individual considers environmentally important" (Stern 2000, p. 416). Therefore, we include an environmental concern index (*Enviro*) as the general atti-

---

[6] As shown in Beaumais et al. (2010), a water habit index was constructed by calculating the mean score on the answers related to the values of water use/conservation habits that were elicited by the survey (possible answers were 1 = yes or 0 = no).

[7] Fortunately, since nowadays LCM routines are available through statistical packages such as Stata, estimating them should be within the reach of most analysts.

tude towards the environment may influence the preferences and, therefore, the membership to a certain water use profile.

The second type of causal variables that we should take into account is personal capabilities, among which we included knowledge and skills that may affect the drivers of residential water demand. In the estimation we considered a binary variable related to the knowledge of the existence of an environmental campaign (*Campaign*). We also included the age (*Age*) of the head of the household, since it has been commonly considered as a determinant of environmentalism and, more specifically, water demand behavior. However, prior literature drew mixed conclusions about this effect. Kantola et al. (1983) and Scott and Willits (1994) found a negative effect on environmental behavior, whereas Gilg and Barr (2006) and Clark and Finley (2007) found that older consumers are more likely to report water conservation intentions.

Automatic processes such as habits and routines may guide behaviors. Therefore, examining the role of habits and routines is fundamental for the analysis of water demand behaviors. As people of different ages may have different habits and routines, we included variables reflecting the proportion of members over 65 years (*Old65*) and those under 16 years (*Young16*). We can expect households with a higher proportion of younger members to have a lower response to changes in the drivers of residential water demand, due to the strong need for more frequent laundering, more frequent showers and use of water-intensive outdoor activities. On the other hand, retired people may also have a lower response, since they are likely to devote more time to activities that involve water use, such as gardening, and spending more time at home. A water-conservation habits index (*Habits*) was also included as a covariate.

The last category of determinants of different water demand behaviors includes contextual factors, such as physical infrastructure and technical facilities, that are also closely related to human behavior. We used a categorical variable that accounts for the number of water efficient electrical appliances in a house (*Electeff*).

In order to control for different consumption patterns that may have been masked by the aggregation, we included as a covariate the standard deviation in water consumption (SD) within a year. This variable is not included in the demand function, because it does not affect directly the level of water consumption but only the water consumption profiles.

Table 4 shows some descriptive statistics and descriptions for the variables included in the LCM.

6 Results

In order to select the model that best fits the data, we estimated several LCMs changing the number of classes and compared likelihood-based model selection criteria, such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), as explained in Sect. 3. The selection criteria, reported in Table 5, lead to different conclusions. The BIC suggests that the 4-class model with variable probabilities fits the distribution better but the AIC suggests that a 5-class model with variable probabilities is best. Therefore, following Nylund et al. (2007), we selected the model that minimizes the BIC, that is, the 4-class model with variable probabilities, since it provides the best fit for these data. The results confirm that the LCM outperforms the OLS model. That is, household heterogeneity is significant and that there seem to be four distinct residential water consumer profiles in Granada for the period 2009–2011, rather than the single one assumed by the conventional OLS approach, which forces all consumers to respond to the same pattern in terms of their water demand.

Table 4 Summary statistics

Variable	Description	Mean	SD	Min.	Max.	N
Dependent variable						
Consumption	Average bimonthly water consumption per year (m^3)	15.746	8.098	1.167	65	3,020
Explanatory variables						
AvP	Average price (€/m^3)	1.699	0.498	0.663	8.076	3,018
AvP2011	*AvPrice* × dummy 2011 (€)	0.696	1.002	0	8.076	3,018
PriceInfo	*AvPrice* × dummy knowledge of tariff structure (€)	0.583	0.852	0	8.076	3,018
HighIncome	=1 if the household income is over 2,700€	0.216	0.411	0	1	3,020
Members	Household size	2.716	1.239	1	9	3,020
Owner	=1 if house is owned by a household member	0.776	0.417	0	1	3,020
Habits	Index of water conservation habits	0.615	0.16	0	1	3,020
Electappl	Number of electrical appliances in the house	1.998	0.608	1	4	3,014
Electeff	Number of efficient water-using electrical appliances in the house	0.812	0.839	0	2	3,020
Covariates[a]						
Enviro	Index of environmental concern ranging from 0 (*not concerned*) to 2 (*very concerned*)	1.820	0.424	0	2	2,903
Campaign	=1 if respondent knows of any water conservation campaign	0.545	0.498	0	1	3,020
Young16	Household members younger than 16 (%)	0.042	0.127	0	0.778	3,020
Old65	Household members over 65 (%)	0.331	0.424	0	1	3,020
Age	Age of household's head	54.361	19.256	18	94	2,968
Std. Dev	Standard deviation in water consumption within a year	2.947	2.688	0.278	44.444	3,020

[a] *Electeff* and *Habits* are also included as covariates

Table 5 Selection criteria for several models

	N	Log-likelihood	Akaike information criterion	Bayesian information criterion
1-Class, constant probability	3,012	−10,225.0	20,472.0	20,538.1
2-Class, constant probability	3,012	−9,509.6	19,069.1	19,219.4
3-Class, constant probability	3,012	−9,235.9	18,547.7	18,776.1
4-Class, constant probability	3,012	−9,170.1	18,442.3	18,748.8
5-Class, constant probability	3,012	−9,139.0	18,405.9	18,790.6
4-Class, variable probability	2,845	−8,072.1	16,294.2	16,740.7
5-Class, variable probability	2,845	−8,116.8	16,425.5	16,996.9

Moreover, as explained in Sect. 3, the distribution of residential water consumption in the city of Granada is asymmetric and, as shown in Fig. 2, the 4-class model fits the data better than both the OLS and the random-effects models.
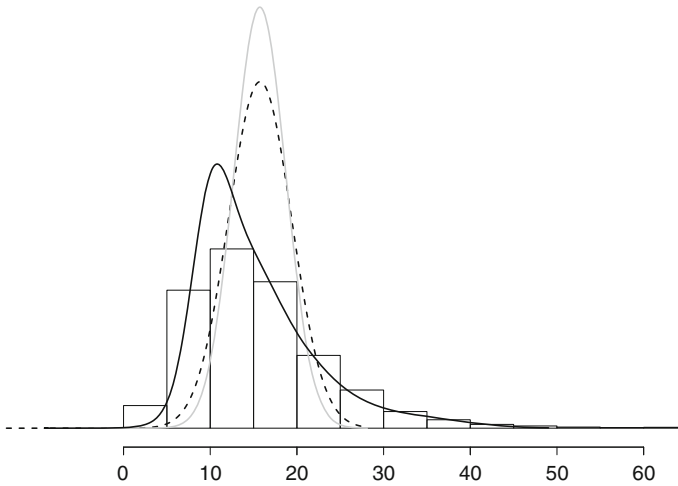
Fig. 2 Distribution of residential water demand in Granada 2009–2011 (*histogram*). Distribution of the predicted value using OLS model (*dashed line*), random-effects model (*grey line*) and the 4-classes LCM with variable probabilities (*black line*)

Table 6 Water consumption statistics by class

Variable	Mean	SD	Min	Max	Classification based on posterior probability (%)
Class 1	9.8910	2.4020	4.1667	16.3333	0.2619
Class 2	33.8012	9.8352	1.1667	65	0.0692
Class 3	13.2829	3.94914	3	22.3333	0.3367
Class 4	18.9621	6.6983	1.333	38.1667	0.3322

Table 6 presents selected descriptive statistics for water consumption by class. On average, the first and the second class are the ones with the lowest and the highest average water consumption respectively. However, as explained in Sect. 3, it should be stressed that consumers are not sorted based on the values taken by the dependent variable (water consumption)[8] but instead according to the similarity of their conditional distributions of the error component, as shown by the minimum and maximum values. It is also worth noting that Class 2, the smallest (representing just over 6% of the sample), has the highest estimated standard deviation in the distribution of consumption. That is, their consumption levels within this class are the most variable. Moreover, Table 7 shows that most of the explanatory variables are not significant. That is, it is not possible to identify a pattern in residential water consumption. As explained by Cameron and Trivedi (2005, p. 625), additional classes may be the result of the LCM grouping outliers. Therefore, considering that this is the smallest class and that we cannot identify the drivers of water demand, we may think that households that belong to this class are in fact outliers whose consumption patterns cannot easily be explained by the usual determinants of the quantity of water demanded.

---

[8] The sample could be split in arbitrary intervals by modeling the inverse of the cumulative distribution function (CDF) of the dependent variable using Quantile Regression.

Table 7 Estimated water demand models

	OLS	RE	LCM			
			Class 1	Class 2	Class 3	Class 4
Main						
AvP	−4.048**	−3.015***	−1.420	18.08	−2.756**	−5.297**
	(−2.32)	(−4.26)	(−0.82)	(0.54)	(−2.46)	(−2.35)
AvP11	0.740*	0.605***	0.354	−5.214	0.557**	0.704
	(1.91)	(4.13)	(0.90)	(−0.65)	(2.29)	(1.50)
Priceinfo	0.488*	0.430*	−0.0475	2.105	0.0731	0.242
	(1.85)	(1.89)	(−0.46)	(1.17)	(0.40)	(0.84)
HighIncome	−0.942*	−0.734	−0.255	−4.263	−0.762***	−0.867
	(−1.88)	(−1.54)	(−1.60)	(−1.42)	(−2.62)	(−1.18)
Members	1.806***	1.907***	0.626***	3.801***	1.090***	1.828***
	(9.08)	(9.59)	(7.07)	(3.44)	(10.09)	(7.27)
Electappl	1.450***	1.403***	0.712***	4.056*	0.854***	0.958*
	(3.53)	(3.59)	(4.13)	(1.85)	(3.79)	(1.69)
Habits	−2.322	−1.677	0.0481	−22.91	−0.385	−3.362
	(−1.58)	(−1.16)	(0.06)	(−1.43)	(−0.24)	(−1.63)
Electeff	−0.467*	−0.572**	0.285***	2.290	0.372	1.114**
	(−1.86)	(−2.27)	(2.61)	(0.69)	(1.59)	(2.45)
Owner	−0.391	−0.486	−0.0409	−3.766	0.0908	−0.720
	(−0.69)	(−0.91)	(−0.21)	(−1.14)	(0.27)	(−1.00)
ControlFunction	−2.323	−0.672	−11.11***	−17.75	−8.578***	−2.297
	(−1.38)	(−1.00)	(−6.18)	(−0.74)	(−6.32)	(−1.12)
Constant	16.34***	14.23***	8.248***	1.148	12.11***	21.97***
	(5.24)	(7.83)	(2.62)	(0.02)	(4.84)	(4.44)
Latent class probability						
Enviro			−0.227	0.0550	0.312	
			(−0.86)	(0.14)	(1.21)	
Campaign			0.0303	0.703**	0.0439	
			(0.13)	(2.15)	(0.22)	
Electeff			0.405*	−0.930**	0.127	
			(1.75)	(−2.35)	(0.92)	
Habits			−0.389	1.049	−1.508*	
			(−0.30)	(0.65)	(−1.70)	
Young16			1.848**	0.456	1.418*	
			(2.02)	(0.34)	(1.96)	
Old65			0.976**	−0.159	0.225	
			(2.19)	(−0.27)	(0.61)	
Age			−0.0234**	0.020	−0.011	
			(−2.44)	(1.30)	(−1.56)	
Std. Dev			−1.805***	0.291***	−0.520***	
			(−11.34)	(5.56)	(−7.14)	
Constant			4.599***	−4.429***	2.136***	
			(4.40)	(−3.15)	(2.69)	

t statistics in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 8 Price elasticities of demand

Model	Price elasticities	
	2009–2010	2011 Effect
Single-class model	−0.4368**	0.0327*
Random-effects model	−0.3255***	0.0268***
4-Class model		
Class 1	−0.2357	0.0224
Class 2	0.9675	−0.1221
Class 3	−0.3419**	0.0300**
Class 4	−0.4951**	0.0264

t statistics in parentheses

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Next, in Table 7, we present the results of the 4-class model with variable probabilities. We also report the results of the OLS model, i.e., a single-class model, and a random-effects model that will be used to assess the importance of household heterogeneity.

First, the estimation of the single-class model and the random-effects model suggests that the demand is price inelastic, being the price elasticity at the mean sample level −0.4368 and −0.3255 respectively as shown in Table 8. However, turning to the 4-class model, we find that for Classes 1 and 2, which include 26.19 and 6.92 % of the observations respectively, price has no significant impact on residential water demand. Therefore, the effect of a change in price would be overestimated for those consumers in Classes 1 and 2 when using the other two models. In contrast, for the remaining classes, price is significant, but price elasticities vary across classes, with the fourth class having the most elastic water demand. This heterogeneity in terms of price elasticities is masked when estimated through a single-class model and a random-effects model. As explained in Sect. 5, in order to allow for the possibility that price elasticity differs between 2009–2010 and 2011, that is, between the period before and after the change in the price structure, we included the interaction involving a dummy variable for 2011 and the lagged average price. Table 8 shows that water demand becomes more inelastic in the single-equation and random-effects estimation. In contrast, the 4-class model suggests that, those consumers in Class 3 have become significantly less responsive to price after the 2011 change in the price structure. That is, the single-class model and random-effects model identify a shift in the demand function for all consumers, while the 4-class model identifies this shift for a specific group of consumers implying that the change in the price structure was not similarly perceived by all consumers. Furthermore, knowledge of the price structure has no significant impact on price elasticity in the 4-class model, but it has a positive and significant effect at the 10 % level both in the single-class model and random-effects model. The lack of significance of this variable may be due to the fact that knowledge of the price structure, that is, that consumers know that the price structure is based on IBP, does not imply that consumers are aware of the block where they consume and the set of marginal prices per block.

Under the single-class model, we find that the high-income indicator has a negative and significant impact on the amount of water consumed. In the 4-class model, this only applies for Class 3. This negative effect may be reflective of water conservation measures resulting from the investment in water-saving devices by this class of consumers. The income variable coefficients corresponding to the remaining classes are not significant. Therefore, a higher

Table 9 Elasticity of water demand with respect to family size

Model	Elasticities
Single-class model	0.3121***
Random-effects model	0.3289***
4-Class model	
Class 1	0.1571***
Class 2	0.3185***
Class 3	0.2257***
Class 4	0.2731***

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

level of income is not associated with a higher demand of water for all the users in the sample, as opposed to what the OLS model would suggest.

The household size has a positive estimated effect on water consumption in all the models. For comparison purposes across models, the values of average elasticities of water demand with respect to household size from the models are presented in Table 9. Overall, the elasticities with respect to family size are quite heterogeneous among classes. However, the results show that, in every case, an increase in water use is less than proportional to an increase in the number of persons per household. This is consistent with other studies that have found economies of scale (Arbués and Villanúa 2006).

Electappl seems to have a positive and significant effect in both the OLS and the random-effects models. That is, the higher the number of electrical- and water-appliances, the higher the level of water consumption. These coefficients are significant for the four classes in the LCM and differ across classes, which suggests that the effect of a higher number of appliances on water demand is heterogeneous.

The variable *Habits* does not have a significant effect on water demand in any of the models that we have estimated. Self-reported water-conservation habits may not have a significant effect on water consumption after controlling for other variables. However, the effect of this variable on water demand may also depend on the habits index used and when those habit are measured, as in Trumbo and O'Keefe (2005) that they measured self-reported behaviors related to water conservation across a two-year time frame, finding that self-reported water-conservation in 1998 had a significant effect on conservation intentions in 2000. That is, future intentions may be affected by past water-conservation habits.

The number of efficient electrical appliances has a negative and significant effect at the 10 % level in both the single-class model and the random-effects model. However, when we estimate the 4-class model, this variable has no significant effect for Classes 2 and 3, although it does have a positive and significant effect for Classes 1 and 4. The installation of water-efficient appliances should reduce water use, although several studies have also found the opposite effect (Campbell et al. 2004; Inman and Jeffrey 2006; Fielding et al. 2012). Among the possible causes of this positive effect on water demand is the possibility of a rebound effect, that is, smaller water savings than expected from the installation of water-efficient equipment due to behavioral changes that partially offset technical efficiency gains, and the fact that investments on water-efficient appliances may be related to activities that imply higher water consumption (Fielding et al. 2012). However, given the unstable behavior and high potential for biasedness[9] affecting the estimates associated with this variable and the

---

[9] The potential endogeneity of habits or efficient electrical appliances is not considered in the paper, since we are not trying to make any type of causal inference about the impact of habits on water consumption. In the terminology of (Angrist and Pischke 2009, p. 68), these variables can be referred to as proxy control variables,

habits index, we make no strong claims about their validity and the resulting conclusions should be viewed with caution.

The coefficient of the binary variable indicating home ownership is not significant in any of the models estimated, indicating that owner occupiers do not differ significantly in terms of their water demand depending on whether they own their home or not. This result could be due to the high proportion of home ownership in Granada (as in the rest of Spain), on the one hand, and the likelihood that many of those who do not own their homes are actually students who might make substantial proportion of their water use (for laundry, etc.) at their family home outside the city, on the other.

The lower portion of Table 7 presents the estimated coefficients of the covariates in the membership function, with Class 4 as the reference category. We can see that most of the covariates do not have statistically significant coefficients. That is, they seem to have very little predictive power about individuals' preferences about water demand. Therefore, if we had divided the sample into explicitly defined groups based on these observable characteristics and self-reported valuations that a priori we expected to be proxies for unobserved preferences about water consumption, we would have misclassified individuals. In this case, LCMs seem to be adequate when some sources of heterogeneity remain unobserved.

However, there are some covariates that provide some information for identifying the different water demand profiles. *Campaign*, *Electeff*, *Habits*, *Young16*, *Old65*, *Age* and *Std. Dev* are all statistically significant. The first class has a positive and significant coefficient for *Young16* and *Old65*. Therefore, the proportion of children under 16 and adults over 65 is expected to be higher than in Class 4. Class 1 is also characterized by younger household heads and a smaller standard deviation in water consumption within the year. This class also has a positive and significant coefficient for *Electeff*, suggesting that consumers in this class have a higher number of efficient electrical appliances than those in Class 4.

Class 2 is characterized by consumers who are aware of the existence of campaigns to promote water savings and who have fewer efficient appliances compared to those in the fourth class. Moreover, consumers in Class 2 have a higher standard deviation in annual water consumption. As explained above, consumers in this group lack a clear pattern of consumption that can be explained by the standard drivers of water demand, not only is the standard deviation in water consumption among consumers in this class is the highest but also the within-year variation for each consumer.

We estimated that Class 3 is characterized by households with a significantly lower score on the water habits index, a significantly higher proportion of children under 15, and a significantly lower standard deviation in annual water consumption relative to the fourth class.

Finally, Class 4, by implication, is defined by households with a lower proportion of children (compared to Classes 1 and 3) and adults over 65 (relative to Class1). However, the head of the household is older than those in the first class. Consumers in this class are less aware of the existence of water-saving campaigns than those in the second class, have fewer water-conservation habits than consumers in the third class, and have a lower number of efficient appliances than consumers in the first class, but higher that consumers in the second class. Regarding the variability (as measured by its standard deviation) of annual

in the sense that they are included in the regression in order to serve as a measure of the observed water behavior and in order to avoid omitted variable bias. Including this variable would not generate a regression coefficient of interest but it may be an improvement over the alternative of using no control. Moreover, we test the robustness of the results by excluding these variables, finding that the other coefficient estimates (not reported but available upon request) remain qualitatively unchanged.

Table 10 Water demand, by Income/SD groups

	(1)	(2)	(3)	(4)
AvP	−5.866***	−0.944	−4.796	−3.925
	(−3.07)	(−0.16)	(−1.64)	(−0.66)
AvP2011	1.255***	−0.262	0.900	0.580
	(2.99)	(−0.20)	(1.36)	(0.44)
Priceinfo	−0.241	0.755	0.994**	0.913
	(−0.97)	(1.31)	(2.41)	(1.19)
Members	0.989***	1.762***	1.684***	1.902***
	(4.85)	(4.55)	(4.77)	(3.76)
Electappl	1.004***	1.137	0.998	2.424*
	(2.87)	(1.26)	(1.43)	(1.80)
Owner	−0.343	0.016	1.198	3.569**
	(−0.63)	(0.01)	(1.13)	(1.99)
Habits	−1.412	−1.204	0.961	−4.729
	(−1.02)	(−0.38)	(0.34)	(−1.13)
Electeff	−0.113	−0.574	−0.076	−0.772
	(−0.43)	(−1.08)	(−0.18)	(−0.92)
ControlFuntion	−1.991	−3.622	−1.193	−2.607
	(−1.05)	(−0.66)	(−0.40)	(−0.50)
Constant	19.30***	15.20	11.86*	13.43
	(5.28)	(1.56)	(1.93)	(1.15)
N	1,491	872	424	225
Log-likelihood	−4,610.9	−3,131.8	−1,279.4	−763.2
AIC	9,241.8	6,283.6	2,578.9	1,546.4
BIC	9,294.9	6,331.3	2,619.4	1,580.5

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

water consumption, households in this class have a larger dispersion than those in Class 1 and 3, but smaller than households in the second class.

6.1 Sensitivity Analysis

We compared the LCM to an alternative model[10] in order to show that the LCM can better capture heterogeneity in the sample. As noted in the introduction, several studies have sorted consumers into different groups based on observable characteristics, such as income. In our own comparative exercise and in order to make groups comparable to those estimated using the LCM, we divided the sample into four sub-groups based on income and the standard deviation [11] and we maintained the same demand specification. As a comparison, and although the LCMs is a probabilistic model, we grouped consumers based upon their estimated modal probability. The results are shown in Table 10.

---

[10] We thank an anonymous referee for suggesting the analysis included in this section.

[11] These groups are defined based on their sample average of the standard deviation in water consumption within a year and the two possible categories of the income variable.

Table 11 Measures of homogeneity and heterogeneity

	Homogeneity	Heterogeneity
Latent Class model	5.0241	0.1091
Grouping based Income/SD	6.5698	0.8548

In the spirit of Nguyen and Rayward-Smith (2008) and Eshghi et al. (2011), we used two measures to evaluate the performance of the different grouping methodologies: the homogeneity of the observations within each group, and the heterogeneity between groups.

To measure the level of homogeneity within groups, we computed the standard deviation of the residuals in each group and then we summed the indicator across groups and divided by the number of groups.

$$s(j) = \sqrt{\frac{\sum_{i=1}^{N(j)} (\epsilon_{ij} - \epsilon_j)^2}{N(j) - 1}} \tag{5}$$

$$S = \sum_{j=1}^{J} \frac{s(j)}{J} \tag{6}$$

To measure the level of heterogeneity between groups, we considered the ratio of the difference between the observed realization of the value of the dependent and the predicted value to the difference between the observed value and the value that would be predicted for that observation if it were assigned to a different group. This indicator was computed for each group, summed across groups, and then divided by the number of groups.

$$\sum_{i=1}^{N(j)} \frac{(y - \hat{y}_j)^2}{(y - \hat{y}_{k \neq j})^2} \tag{7}$$

$$H = \sum_{j=1}^{J} \frac{h(j)}{J} \tag{8}$$

Table 11 shows the measures of homogeneity and heterogeneity calculated using the results from each methodology. These results support the conclusion that our LCM estimation provides the most homogeneous groups, while it succeeds in differentiating among these groups.

7 Conclusions

This study provides strong evidence of unobserved heterogeneity in residential water demand in the city of Granada for the period 2009–2011. We identified four different residential water consumer profiles in Granada for the period 2009–2011 using a LCM, rather than the common profile assumed by single equation approaches, and this estimation allowed us to observe four distinct prices responses. Moreover, our sensitivity analysis shows that the LCM technique is an appropriate method to group observations homogeneously.

Water demand is found to be perfectly inelastic for two of the classes we identified. The proportion of consumers who belong to these classes, based on the posterior probabilities, exceeds 33 %. The effect of a change in price would be overestimated for these two groups of consumers, which represent an substantial proportion of the sample, if price elasticities were

estimated using a single-class model. The implementation of pricing policies would likely be less effective in reducing water consumption for these two groups of consumers in the future. Particularly, one of the groups that are relatively insensitive to price changes (which represents a fourth of the households) registers low average consumption levels. This result may be in line with those obtained estimating residential water demand functions based on a Stone–Geary utility function (Gaudin et al. 2001; Martínez-Espiñeira and Nauges 2004; Dharmaratna and Harris 2012; Garcia-Valiñas et al. 2014), which suggest the existence of a nondiscretionay amount of water that is not sensitive to price changes and an additional quantity devoted to discretionary uses that does respond to price variations.

Identifying different price elasticities allows regulators to predict more accurately the effect of different water conservation policies. Our analysis suggests that the focus of a water demand management policy could be tailored to the specific demand function of a particular group of consumers. Indeed, in order to reduce water consumption, pricing and non-pricing policies, such as education programs, water rationing, retrofit subsidies or public information campaigns can be jointly applied to the most price-responsive groups of consumers. However, non-pricing policies should be intensified in the case of the least price-responsive consumers, especially for the class that has a low level of water consumption, likely to be mainly nondiscretionary and, therefore, hard to adjust in the short-run. Promoting water-saving habits and the installation of efficient appliances could be useful to reduce both discretionary and non-discretionary residential water consumption (Garcia-Valiñas et al. 2014). We illustrate how the analysis of membership probabilities makes it possible to identify the characteristics of the users that belong to a given class, which should make it easier to tailor water conservation programs to best suit the response patterns of different user groups.

We have shown that the use of Latent Class Analysis shows a reasonable degree of potential as a tool to improve our understanding of residential water demand. This is, however, the first time that, to our knowledge, the use of Latent Class Analysis has been used in the estimation of water demand functions. It would, therefore, be interesting to replicate this type of work using similar data from other jurisdictions to find out whether and to which extent our results can be generalized further.

Appendix 1

As explained in Sect. 5, the estimation using LCMs is nonlinear, therefore, two-stage least squares models are likely to be inconsistent (Howard and Roe 2013). Therefore, we used a control function approach to correct for price endogeneity.

Consider the model:

$$WD = \delta z_1 + \alpha AvgP + u_1 \tag{9}$$

where WD is residential water demand, $AvgP$ the average price (i.e., the endogenous explanatory variable) and z_1 a vector of exogenous explanatory variables.

This methodology uses the same first stage that would be used in 2SLS, that is, the endogenous explanatory variable is regressed on the exogenous explanatory variables and the set of instruments. Let z_2 denote a vector of instruments.

$$AvgP = \pi_1 z_1 + \pi_2 z_2 + u_2$$
$$E(z_1' u_2) = 0 \quad E(z_2' u_2) = 0 \tag{10}$$

The average price would be endogenous if and only if u_1 is correlated with u_2, being $\gamma = E(u_2 u_1)/E(u_2^2)$.

$$u_1 = \gamma u_2 + \epsilon \tag{11}$$

Since u_1 and u_2 are uncorrelated with z_1 and z_2, $E(u_2 \epsilon) = 0$ and $E(z_2 \epsilon) = 0$. Substituting Eq. (11) into Eq. (9):

$$WD = \delta z_1 + \alpha AvgP + \gamma u_2 + \epsilon \tag{12}$$

In this equation u_2 is included as an explanatory variable. As explained above, ϵ is uncorrelated with u_2, z_1 and z_2. Moreover, $AvgP$ is defined as a linear function of the explanatory variables, the instruments and the residual u_2, so $AvgP$ is uncorrelated with ϵ and therefore, δ and α can be consistently estimated with Eq. (12).

Results from the first stage estimation as defined in Eq. (10) are presented in Table 12. Following Hewitt and Hanemann (1995) and Olmstead (2009b), we used the full set of marginal prices in each block as instruments. The Hansen test indicated that all the instruments are exogenous. The residuals were obtained from this estimation and then, the second stage estimation included the average price and the control function as defined in Eq. (11), that is, the residuals from the first stage and a standard normal random variable (Howard and Roe 2013).

Appendix 2

Selected questions from the survey used to construct a water habits index

P.17. In general, do you have any of the following water conservation habits in the household?

(a) Do you recycle water, for example, making use of the water while you wait for the shower to get hot?
(b) Do you store drinking water in the refrigerator rather than letting the tap run every time you want a cool glass of water?
(c) Do you defrost food in advance in order to avoid using running hot water to thaw meat or other frozen foods?
(d) Do you fill the sink with water when washing dishes by hand?
(e) Do you operate automatic dishwashers and washing machines only when they are fully loaded?
(f) Do you slightly turn off the backflow valve to reduce the tap flow?
(g) Do you use a rubbish bin in the toilet rather than flushing the toilet unnecessarily?
(h) Do you avoid letting water run while brushing your teeth?
(i) Do you take shorter showers?
(j) Do you avoid washing the cars with drinking water?

Table 12 Control function estimation

	AvP
Block 1	−1.035
	(−0.60)
Block 2	29.948***
	(4.47)
Block 3	−26.959***
	(−4.06)
Block4	18.151***
	(4.12)
Block 5	−12.443***
	(−4.07)
Highincome	0.038
	(1.22)
Members	−0.039***
	(−4.40)
Electappl	−0.078***
	(−3.76)
Habits	−0.028
	(−0.29)
Electeff	−0.014
	(−1.05)
Owner	0.049**
	(2.04)
Year 2011	−34.596***
	(−4.16)
Priceinfo	−0.012
	(−0.53)
Constant	20.368***
	(3.99)
N	3,012
Hansen test (over identification	5.961
test of all instruments)	$\chi^2(4)$ $p$value $= 0.2021$
Endogeneity test of endogenous regressors	17.670
	$\chi^2(1)$ $p$value $= 0.0000$

t Statistics in parentheses
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

References

Agthe DE, Billings RB, Dobra JL, Raffiee K (1986) A simultaneous equation demand model for block rates. Water Resour Res 22(1):1–4

Angrist JD, Pischke JS (2009) Mostly harmless econometrics: an empiricist's companion, Chapter 3. Princeton University Press, Princeton

Arbués F, García-Valiñas M, Martinez-Espiñeira R (2003) Estimation of residential water demand: a state-of-the-art review. J Socio-Econ 32:81–102

Arbués F, Villanúa I (2006) Potential for pricing policies in water resource management: Estimation of urban residential water demand in Zaragoza, Spain. Urban Stud 43(13):2421–2442

Ayyagari P, Deb P, Fletcher J, Gallo W, Sindelar J (2013) Understanding heterogeneity in price elasticities in the demand for alcohol for older individuals. Health Econ 22(1):89–105

Beaumais O, Briand A, Millock K, Nauges C (2010) What are households willing to pay for better tap water quality? A cross-country valuation study. Université Paris 1 Panthéon-Sorbonne (Post-Print and Working Papers), HAL

Blundell R, Powell JL (2003) Endogeneity in nonparametric and semiparametric regression models. Econom Soc Monogr 36:312–357

Bockstael NE, McConnell KE (1983) Welfare measurement in the household production framework. Am Econ Rev 73(4):806–814

Boter J, Rouwendal J, Wedel M (2005) Employing travel time to compare the value of competing cultural organizations. J Cult Econ 29:19–33

Cameron AC, Trivedi PK (2005) Microeconometrics: methods and applications. Cambridge University Press, New York

Campbell D, Hensher DA, Scarpa R (2011) Non-attendance to attributes in environmental choice analysis: a latent class specification. J Environ Plan Manag 54(8):1061–1076

Campbell HE, Johnson RM, Larson EH (2004) Prices, devices, people, or rules: the relative effectiveness of policy instruments in water conservation. Rev Policy Res 21(5):637–662

Clark WA, Finley JC (2007) Determinants of water conservation intention in blagoevgrad, bulgaria. Soc Nat Resour 20(7):613–627

Coleman EA (2009) A comparison of demand-side water management strategies using disaggregate data. Public Works Manag Policy 13(3):215–223

d' Uva TB (2006) Latent class models for utilisation of health care. Health Econ 15:329–343

Dalhuisen JM, Florax R, Groot HD, Nijkamp P (2003) Price and income elasticities of residential water demand: a meta-analysis. Land Econ 79(2):292–308

Deb P, Trivedi PK (2002) The structure of demand for health care: latent class versus two-part models. J Health Econ 21:601–625

Dharmaratna D, Harris E (2012) Estimating residential water demand using the stone-geary functional form: the case of Sri Lanka. Water Resour Manag 26(8):2283–2299

Eshghi A, Haughton D, Legrand P, Skaletsky M, Woolford S (2011) Identifying groups: a comparison of methodologies. J Data Sci 9:271–291

Fernandez-Blanco V, Orea L, Prieto-Rodriguez J (2009) Analyzing consumers heterogeneity and self-reported tastes: an approach consistent with the consumer's decision making process. J Econ Psychol 30:622–633

Fielding KS, Russell S, Spinks A, Mankad A (2012) Determinants of household water conservation: the role of demographic, infrastructure, behavior, and psychosocial variables. Water Resour Res 48(W10510). doi:10.1029/2012WR012398

García-Valiñas MA (2005) Efficiency and equity in natural resource pricing: a proposal for urban water distribution services. Environ Resour Econ 32(3):183–204

Garcia-Valiñas MA, Athukorala W, Wilson C, Torgler B, Gifford R (2014) Nondiscretionary residential water use: the impact of habits and water-efficient technologies. Aust J Agric Resour Econ 58(2):185–204

Gaudin S (2006) Effect of price information on residential water demand. Appl Econ 38(4):383–393

Gaudin S, Griffin RC, Sickles RC (2001) Demand specification for municipal water management: evaluation of the Stone-Geary form. Land Econ 77(3):399–422

Gilg A, Barr S (2006) Behavioural attitudes towards water saving? Evidence from a study of environmental actions. Ecol Econ 57(3):400–414

Grafton RQ, Ward MB, To H, KT T (2011) Determinants of residential water consumption: evidence and analysis from a 10-country household survey. Water Resour Res 47(W08537). doi:10.1029/2010WR009685

Greene W, Hensher D (2013) Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. Appl Econ 45(14):1897–1902

Grisolía JM, Willis KG (2012) A latent class model of theatre demand. J Cult Econ 36:113–139

Hensher DA, Greene WH (2003) The mixed logit model: the state of practice. Transportation 30:133–176

Hess S, Ben-Akiva M, Walker J (2011) Advantages of latent class over continuous mixture of logit models. Working paper, University Press, Harrisburg

Hewitt JA, Hanemann WM (1995) A discrete/continuous choice approach to residential water demand under block rate pricing. Land Econ 71:173–192

Howard G, Roe BE (2013) Stripping because you want to versus stripping because the money is good: a latent class analysis of farmer preferences regarding filter strip programs. In: 2013 annual meeting, August 4–6, 2013, Washington, D.C., Agricultural and Applied Economics Association

Hyppolite J, Trivedi P (2012) Alternative approaches for econometric analysis of panel count data using dynamic latent class models (with application to doctor visits data). Health Econ 21(1):101–128

Imbens GM, Wooldridge JM (2007, Summer) Control function and related methods. The National Bureau of Economic Research (NBER). http://users.nber.org/~confer/2007/si2007/WNE/lect_6_controlfuncs.pdf

Inman D, Jeffrey P (2006) A review of residential water conservation tool performance and influences on implementation effectiveness. Urban Water J 3(3):127–143

Kantola SJ, Syme GJ, Nesdale AR (1983) The effects of appraised severity and efficacy in promoting water conservation: an informational analysis. J Appl Soc Psychol 13(2):164–182

Krause K (2003) The demand for water: consumer response to scarcity. J Regul Econ 23(2):167–191

Lopez-Gunn E, Zorrilla P, Prieto F, Llamas M (2012) Lost in translation? water efficiency in spanish agriculture. Agric Water Manag 108:83–95

Mansur ET, Olmstead SM (2012) The value of scarce water: measuring the inefficiency of municipal regulations. J Urban Econ 71:332–346

Martínez-Espiñeira R, Nauges C (2004) Is all domestic water consumption sensitive to price control? Appl Econ 36(15):1697–1703

Miyawaki K, Omori Y, Hibiki A (2010) Panel data analysis of japanese residential water demand using a discrete/continuous choice approach. Global COE Hi-stat discussion paper series gd09-123, Institute of Economic Research, Hitotsubashi University

Nataraj S, Hanemann MW (2011) Does marginal price matter? A regression discontinuity approach to estimating water demand. J Environ Econ Manag 61(2):198–212

Nauges C, Whittington D (2010) Estimation of water demand in developing countries: an overview. World Bank Res Obs 25(2):263–294

Nguyen QH, Rayward-Smith VJ (2008) Internal quality measures for clustering in metric spaces. Int J Bus Intell Data Min 3(1):4–29

Nieswiadomy ML, Molina DJ (1988) Urban water demand estimates under increasing block rates. Growth Change 19(1):1–12

Nieswiadomy ML, Molina DJ (1989) Comparing residential water demand estimates under decreasing and increasing block rates using household data. Land Econ 65(3):280–289

Nylund KL, Asparouhov T, Muthén BO (2007) Deciding on the number of classes in latent class analysis and growth mixture modeling: a monte carlo simulation study. Struct Equ Model Multidiscip J 14(4):535–569

Olmstead SM (2009b) Reduced-form versus structural models of water demand under nonlinear prices. J Bus Econ Stat 27(1):84–94

Olmstead SM, Stavins RN (2007) Managing Water Demand. Price vs. Non-price conservation programs. A Pioneer Institute Water paper no 39

Patunru AA, Braden John B, Chattopadhyay S (2007) Who cares about environmental stigmas and does it matter? A latent segmentation analysis of stated preferences for real estate. Am J Agric Econ 89(3):712–726

Pint E (1999) Household responses to increased water rates during the California drought. Land Econ 75:246–266

Polebitski A, Palmer R (2010) Seasonal residential water demand forecasting for census tracts. J Water Resour Plan Manag 136(1):27–36

Renwick M, Archibald S (1998) Demand side management policies for residential water use: Who bears the conservation burden? Land Econ 74:343–359

Renwick ME, Green RD (2000) Do residential water demand side management policies measure up? An analysis of eight California water agencies. J Environ Econ Manag 40:37–55

Roibás D, García-Valiñas MA, Wall A (2007) Measuring welfare losses from interruption and pricing as responses to water shortages: an application to the case of Seville. Environ Resour Econ 38(2):231–243

Ruijs A, Zimmermann A, Berg AMvd (2008) Demand and distributional effects of water pricing policies. Ecol Econ 66(2–3):506–516

Russell S, Fielding K (2010) Water demand management research: a psychological perspective. Water Resour Res 46(W05302). doi:10.1029/2009WR008408

Scarpa R, Thiene M, Tempesta T (2007) Latent class count models of total visitation demand: days out hiking in the eastern Alps. Environ Resour Econ 38(4):447–460

Scarpa R, Willis KG, Acutt M (2005) Individual-specific welfare measures for public goods: a latent class approach to residential customers of Yorkshire Water. Econom Inf Nat Resour Manag 14:316–337

Scott D, Willits FK (1994) Environmental attitudes and behavior: a Pennsylvania survey. Environ Behav 26(2):239–260

Shen J (2010) Latent class model or mixed logit model? A comparison by transport mode choice data. Appl Econ 41:2915–2924

Shen J, Sakata Y, Hashimoto Y (2006) A comparison between latent class model and mixed logit model for transport mode choice: evidences from two datasets of Japan. Osaka University, Graduate School of Economics and Osaka School of International Public Policy (OSIPP). Discussion Papers in Economics and Business (06-05)

Stern PC (2000) New environmental theories: toward a coherent theory of environmentally significant behavior. J Soc Issues 56:407–424

Stigler GJ, Becker GS (1977) De gustibus non est disputandum. Am Econ Rev 67(2):76–90

Strong A, Smith VK (2010) Reconsidering the economics of demand analysis with kinked budget constraints. Land Econ 86(1):173–190

The World Bank (2012) Turn down the heat. Why a 4 C warmer world must be avoided. Technical report, The World Bank

Trumbo CW, O'Keefe GJ (2005) Intention to conserve water: environmental values, reasoned action, and information effects across time. Soc Nat Resour 18(6):573–585

UNDESA (2009) World population prospects, the 2008 revision—executive summary. Technical report, UNDESA (United Nations Department of Economic and Social Affairs)

Worthington AC, Higgs H, Hoffmann M (2009) Residential water demand modeling in Queensland, Australia: a comparative panel data approach. Water Policy 11(4):427–441

Worthington AC, Hoffman M (2008) An empirical survey of residential water demand modelling. J Econ Surv 22(5):842–871