XVI
World Water Congress
International Water Resources Association (IWRA)
Cancun, Quintana Roo. Mexico. 29 May- 3 June, 2017

Functional Forms and Price Elasticities in the Discrete-Continuous Choice Model for Residential Water Demand

Abstract

During recent decades, water demand estimation has gained a lot of attention from scholars. From an econometric perspective, the most used functional forms include both the log–log and the linear specifications. Despite the advances in this field, as well as the relevance for policy making, little attention has been paid to the functional form used in these estimations, as most authors have not provided any justification for the chosen functional forms. In this paper, we estimate a discrete continuous choice model for residential water demand using four functional forms (log–log, semi-log, linear, and Stone–Geary) comparing both the expected consumption and the price elasticity. From a policy perspective, our results shed light on the relevance of the chosen function form, for both the expected consumption and the price elasticity.

1 Introduction

In this paper, we estimate a discrete continuous choice (DCC) model for residential water demand using four functional forms, comparing both the expected consumption and the price elasticity. *Arbués et al.* [2003], *Dalhuisen et al.* [2003] and *Ferrara* [2008] show several functional forms used in the literature to specify the water demand equation, they argue that the selection of a functional form may affect the estimates of price and income elasticities. Comparison among functional forms is scarce in the literature, and authors generally do not provide any justification for the chosen functional forms. Other studies attempt to mitigate this uncertainty by estimating several functional forms, expanding the number of results available for researchers and policy makers regarding consumption prediction and elasticities.

To the best of our knowledge, an evaluation of the impact of functional forms in the context of a DCC model is not available. Increasing block tariff (IBT) schemes are very common in the literature. For instance, more than 40% of the studies reported by *Dalhuisen* *et al.* [2003] show multiple or non-linear tariffs, while 74% of water utilities in developing countries use an IBT, according to *Fuente et al.* [2016]. Thus, the understanding of the role played by the functional form is key to informing policy makers in the development of water policies.

When dealing with an increasing (or decreasing) price scheme (increasing block tariff, ITF) researchers have to solve the simultaneous choice of both the marginal price and the consumption level. *Hewitt and Hanemann* [1995] and *Olmstead et al.* [2007] solve this simultaneity issue (endogeneity) by using a discrete-continuous choice model to estimate the water demand. They illustrate their solution using a "log-log" (logarithmic or double log) demand equation. The log–log is the prevailing functional form in DCC model literature. This may be because of the difficulties in building the likelihood function in the DCC model, and the even more intricate calculation of price and income elasticities—see *Olmstead et al.* [2007]—or the fact that no software package includes the DCC model. We filled this gap by building the likelihood function for each functional form; we also derive the formula for the expected value and the price elasticity in each case.

Section 2 briefly reviews the literature and presents the functional forms; section 3 shows the DCC choice model and the mathematical expressions of expected consumption and price elasticity for each functional form; Section 4 shows results and hypothesis testing; and section 5 presents the conclusions.

2 Water Demand and Functional Forms

Since *Headley* [1963] and *Howe and Linaweaver* [1967], there has been an increasing number of studies analyzing the factors that influence water consumption, the impact of socio-economic variables on water demand, and the calculation of price and income elasticities. *Headley* [1963] carries out one of the first studies analyzing the impact of income on water consumption, *Howe and Linaweaver* [1967] and *Wong* [1972] include prices as a determinant of household water consumption, while *Wong* [1972] also incorporates the effect of climate variables into the demand equation. But it is in *Young* [1973] where the idea of "water price elasticity" is coined in the literature.

These studies, and the following development of the literature, are well presented in *Arbués et al.* [2003] and *Ferrara* [2008]; both authors identify the functional forms of the demand equation as key in determining the results reported in the literature. *Arbués et al.* [2003] and *Dalhuisen et al.* [2003] present three frequently used functional forms: the linear functional form, the log-log form, and the semi-log form (see Table 1).

In Table 1 w is water consumption in m^3, Z is a vector of sociodemographic and climate variables, P is price, Y is income, μ is a stochastic component and δ, α, and γ are parameters to be estimated. These three functional forms cover more than 95% of all studies on water demand; an important number of those studies use more than one functional form.

Al-Qunaibet and Johnston [1985] and *Gaudin et al.* [2001] also use an alternative functional form known as the Stone–Geary (SG) function, which considers the existence of a minimum level of consumption (subsistence level) in its structure.

Table 1: Functional forms commonly used in the residential water demand

	Equation	Dalhuisen et al. (2003)	Arbués et al. (2003)
Log-log	$\ln w = Z\delta + \alpha \ln P + \gamma \ln Y + \mu$	28	16
Semi-log	$\ln w = Z\delta + \alpha P + \gamma Y + \mu$	3	8
Linear	$w = Z\delta + \alpha P + \gamma Y + \mu$	24	29
SG	$w = Z\delta + \alpha(Y/P) + \gamma(1/P) + u$	3	2
More than one		10 (26.3%)	11 (35.4%)

The vast majority of studies do not provide any justification for the functional form they use in their demand equation [*Arbués et al.*, 2003]. The empirical evidence shows that both the price elasticities and the expected consumption are sensitive to the functional form specification. Thus, it is advisable to estimate several functional forms, providing a range of estimates, to address the uncertainty about the functional form adequacy.

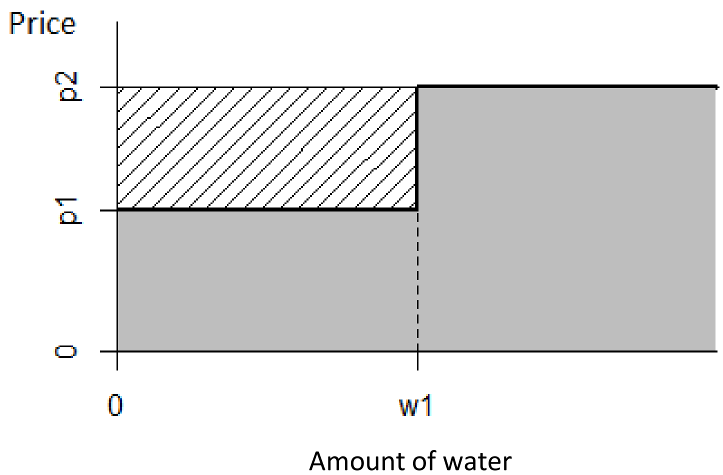
Theoretically, there are some reasons to choose one functional form over others. For instance, water is an essential good, and, therefore, the functional form should consider this fact. Although this rules out the linear functional form, the linear functional form is one of the most common forms used in the literature. On the contrary, both the log–log and semi-log functional forms are asymptotic to zero, showing that water is an essential good, while the Stone–Geary includes a minimum (subsistence) level of consumption in its structure. On the other hand, estimation and calculation of price elasticity are also straightforward in the log–log, semi-log, and linear functional forms, but these desirable attributes are lost in the DCC model.

Furthermore, goodness of fit is an empirical issue; therefore, researchers could estimate several functional forms and select the one with the best fit to the data. Even if we do not want to select a particular functional form, providing estimations from different functional forms enrich the analysis for both researchers and policy makers. Capturing a wider range of possible results will tell us whether the estimates are robust to functional form or, on the contrary, tell us that the variability of the estimates could indicate either poor data or an incorrect estimation strategy.

3 Increasing Block Tariff and Simultaneity: The Discrete-Continuous Choice Models

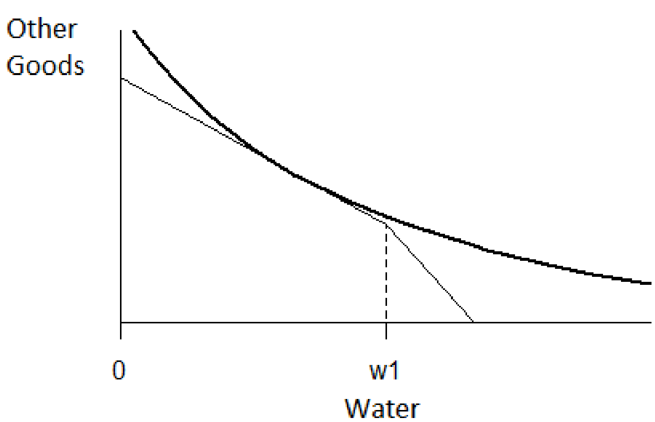
The presence of non-linear price structures, as in the case of an IBT (Figure 1), produces an endogeneity problem that will impose challenges for the estimation of the demand equation. In Figure 1 people can consume either below w_1 or above this threshold (kink point); for quantities below w_1 the consumer will pay p_1, while, for quantities above w_1, he/she will pay p_2. The shaded area represents a “virtual subsidy,” because people consuming above w_1 pay less for the first w_1 m^3.

Figure 1: Example of a two-block increasing price system.



Taylor [1975] shows that a non-linear price system transforms the linear budget constraint in the consumer utility maximization problem into a non-linear budget constraint (in some cases, non-convex), as is shown in Figure 2. The optimal level of consumption can be below, on, or above w_1, and the "virtual subsidy" changes the budget constraint of the consumer. *Nordin* [1976] suggests considering this subsidy by recalculating the household income through adding the virtual income, which is determined as the price difference times w_1.

Figure 2: Effect of a non-linear price system on the budget constraint.



This suggestion solves the fact that people face a nonlinear budget constraint, but does not solve the endogeneity issue, that is, the simultaneity between the consumption level and the price choice. *Hewitt and Hanemann* [1995] suggest the DCC model, using a log–log functional form, to deal with the endogeneity issue, while *Olmstead et al.* [2007] propose an analytical expression for the estimation of the price elasticity within the DCC model. Defining the elasticity in an IBT system is not straightforward, since there are multiple elasticities that can be estimated. For instance, we could estimate the price elasticity for a proportional change in the first price (p_1), the elasticity of a change in the second price (p_2), or the elasticity of a change in any price included in the tariff system. Furthermore, we could be interested in the elasticity associated with a proportional change in the whole price structure.

3.1 The DCC Model: Theoretical and econometric model

Although this model is already presented in several papers [*Hewitt and Hanemann*, 1995; *Moffitt*, 1986; 1989; *Olmstead et al.*, 2007], we repeat some equations of the DCC model to assure the document is self-contained. For our simple two-price case, there are two prices related to each of the tiers (p_1 and p_2), as well as one kink point w_1, which separates block 1 from block 2. This is generalized to k tiers, k prices, and $k-1$ kink points.

The conditional demand represents the water consumption decision made by a consumer, given that he is in a determined consumption block. The conditional demand to a k consumption block is equal to the demand equation evaluated in the marginal price for the corresponding block (p_k), and the household income plus the compensation to the income proposed by Nordin (1976) (d_k) is defined as:

$$d_k = \begin{cases} 0 & si\ k = 1, \\ \sum_{j=1}^{k-1} (p_{j+1} - p_j) w_k & si\ k > 1. \end{cases}$$

The unconditional demand is a function of all consumption blocks and kink points; consequently, it captures the full decision made by the consumer. For instance, the conditional demand under the log–log functional form is:

$$w(p_k, y + d_k) = exp(Z\delta) {p_k}^{\alpha} (y + d_k)^{\gamma} \tag{1}$$

Whereas, under the semi-log form, the conditional demand is:

$$w(p_k, y + d_k) = exp(Z\delta + \alpha p_k + \gamma(y + d_k)) \quad (2)$$

The unconditional demand related to equations (1) and (2) for the simple case of $k = 2$ is:

$$lnw = \begin{cases} lnw_1^* & si\ lnw_1^* < lnw_1 \\ lnw_1 & si\ lnw_1 < lnw_1^*\ y\ lnw_1 > lnw_2^* \\ lnw_2^* & si\ lnw_2^* > lnw_1 \end{cases} \quad (3)$$

w represents the observed water consumption, $w_k^* = w_k^*(Z, p_k, (y + d_k); \alpha, \gamma, \delta)$ is the optimum water consumption in the k block, and w_1 is the kink point.

Equation (3) represents the theoretical model, which is unknown to the researcher. The econometric model incorporates two error terms, following *Burtless and Hausman* [1978], *Moffitt* [1986], and *Hewitt and Hanemann* [1995]: η, which captures the heterogeneity among households, which is not captured by the sociodemographic and climate variables Z; and ε, which represents characteristics that are not observed by either the researcher or the households [*Olmstead et al.*, 2007]. It is assumed that η and ε are independent and normally distributed, with means equal to zero and variances σ_η^2 and σ_ε^2, respectively.

Considering the aforementioned, the unconditional demand is equal to:

$$lnw = \begin{cases} lnw_1^* + \eta + \varepsilon & si\ -\infty < \eta < lnw_1 - lnw_1^* \\ lnw_1 + \varepsilon & si\ lnw_1 - lnw_1^* < \eta < lnw_1 - lnw_2^* \\ lnw_2^* + \eta + \varepsilon & si\ lnw_1 - lnw_2^* < \eta < \infty \end{cases} \quad (4)$$

When using the linear functional form the conditional demand to a k block is equal to:

$$w(p_k, y + d_k) = Z\delta + \alpha p_k + \gamma(y + d_k) \quad (5)$$

Whereas, under the SG the conditional demand is:

$$w(p_k, y + d_k) = Z\delta + \alpha\left(\frac{(y + d_k)}{p_k}\right) + \gamma\left(\frac{1}{p_k}\right) \tag{6}$$

Since they do not have a logarithmic transformation, the unconditional demand is a modification of the equation in both cases (3):

$$w = \begin{cases} w_1^* & si\ w_1^* < w_1 \\ w_1 & si\ w_1 < w_1^*\ y\ w_1 > w_2^* \\ w_2^* & si\ w_2^* > w_1 \end{cases} \tag{7}$$

In addition, the econometric model related to (7) incorporates the heterogeneity errors of households (η) and stochastic (ε):

$$w = \begin{cases} w_1^* + \eta + \varepsilon & si\ -\infty < \eta < w_1 - w_1^* \\ w_1 + \varepsilon & si\ w_1 - w_1^* < \eta < w_1 - w_2^* \\ w_2^* + \eta + \varepsilon & si\ w_1 - w_2^* < \eta < \infty \end{cases} \tag{8}$$

The incorporation of the error terms allows the estimation of equations (4) and (8) through maximum likelihood.

3.2 Estimation

Following *Hewitt and Hanemann* [1995] and *Olmstead et al.* [2007], the likelihood function related to the equation (4) is equal to (see appendix for details):

$$lnL = \sum ln \left[\begin{array}{c} \left(\frac{1}{\sqrt{2\pi}} \frac{exp(-s_1^{*2}/2)}{\sigma_v} (\Phi(r_1^*))\right) \\ + \left(\frac{1}{\sqrt{2\pi}} \frac{exp(-s_2^{*2}/2)}{\sigma_v} (1 - \Phi(r_1^*))\right) \\ + \left(\frac{1}{\sqrt{2\pi}} \frac{exp(-u_1^{*2}/2)}{\sigma_\varepsilon} (\Phi(t_2^*) - \Phi(t_1^*))\right) \end{array} \right]$$

Where:

$$\rho = corr(\varepsilon + \eta, \eta); \qquad v = \eta + \varepsilon$$
$$s_k^* = (lnw_i - lnw_k^*(\cdot))/\sigma_v; \quad u_k^* = (lnw_i - lnw_k)/\sigma_\varepsilon$$
$$t_k^* = (lnw_1 - lnw_k^*(\cdot))/\sigma_\eta; \quad r_k^* = (t_k^* - \rho s_k^*)/\sqrt{1-\rho^2}$$

Our own calculations, based on the initial work by *Moffitt* [1986] and *Moffitt* [1989], show us that likelihood function for the linear and SG functional forms is similar, with the following modification:

$$s_k = (w_i - w_k)/\sigma_v; \qquad u_k = (w_i - w_k)/\sigma_\varepsilon$$
$$t_k = (w_1 - w_k^*(\cdot))/\sigma_\eta; \quad r_k = (t_k - \rho s_k)/\sqrt{1-\rho^2}$$

3.3 Expected value and elasticities

The main challenge in the DCC model is the estimation of the expected value and the elasticities. Since the model captures two decisions, the discrete and continuous decisions, the expected consumption is the sum of the consumption at each tier and kink point, weighted by the probability of being in each tier or kink point. In other words, the expected consumption depends on all prices and kink points, and not solely on the current consumption level. Consumption is the result of a discrete choice among different tiers and, therefore, the expected value needs to capture the stochastic nature of that choice.

For the log–log and semi-log functional forms, it can be shown that that the conditional consumption is:

$$w_k^*(p_k, y + d_k) = exp(Z\delta){p_k}^{\alpha}(y + d_k)^{\gamma} exp(\eta) exp(\varepsilon)$$

$$w_k^*(p_k, y + d_k) = exp(Z\delta + \alpha p_k + \gamma(y + d_k)) exp(\eta) exp(\varepsilon)$$

Since both η and ε are normally distributed, $exp(\eta)$ and $exp(\varepsilon)$ are distributed lognormal. [*Olmstead et al.*, 2007] show that the expected value for the first functional form for the simple case of two tiers is:

$$E(W) = e^{\sigma_\eta^2/2} e^{\sigma_\varepsilon^2/2} \left(w_1^*(p_1, y + d_1) * \pi_1^* + w_2^*(p_2, y + d_2) * \pi_2^* \right) + e^{\sigma_\varepsilon^2/2} w_1 * \lambda_1^* \qquad (9)$$

With

$$\pi_1^* = \Phi\left(\frac{ln(w_1/w_1^*)}{\sigma_\eta} - \sigma_\eta\right)$$
$$\pi_2^* = 1 - \Phi\left(\frac{ln(w_1/w_2^*)}{\sigma_\eta} - \sigma_\eta\right)$$
$$\lambda_1^* = \Phi\left(\frac{ln(w_1/w_2^*)}{\sigma_\eta}\right) - \Phi\left(\frac{ln(w_1/w_1^*)}{\sigma_\eta}\right)$$

Calculating the elasticity is a little more complex. *Hewitt and Hanemann* [1995] calculate the elasticity, simulating a change in 1% of all prices and recalculating the expected value. Olmstead et al. (2007) formalize this approach, developing an analytical expression for the price elasticity as the change in the expected value after a change in a proportion θ in the price vector. They show that, for the log–log functional form the elasticity is:

$$\frac{\partial E(W)}{\partial \theta}\frac{1}{E(W)} = \left(\begin{array}{c} \alpha\left(w_1^* \psi_1 + w_2^* \psi_2 + w_1(\chi_1 - \chi_2)\right) \\ + \gamma\left(d_2\left(\frac{w_2^*}{y + d_2}\right)\left(\psi_2 - \left(\frac{w_1}{w_2^*}\right)\chi_2\right)\right) \end{array} \right) \Big/ \Omega \qquad (10)$$

In which:

$$\psi_1 = \pi_1^* - \frac{1}{\sigma_\eta}\phi\left(\frac{ln(w_1/w_1^*)}{\sigma_\eta} - \sigma_\eta\right)$$
$$\psi_2 = \pi_2^* + \frac{1}{\sigma_\eta}\phi\left(\frac{ln(w_1/w_2^*)}{\sigma_\eta} - \sigma_\eta\right)$$
$$\chi_1 = \frac{1}{\sigma_\eta * e^{\sigma_\eta^2/2}}\phi\left(\frac{ln(w_1/w_1^*)}{\sigma_\eta}\right)$$
$$\chi_2 = \frac{1}{\sigma_\eta * e^{\sigma_\eta^2/2}}\phi\left(\frac{ln(w_1/w_2^*)}{\sigma_\eta}\right)$$
$$\Omega = w_1^*(p_1, y + d_1) * \pi_1^* + w_2^*(p_2, y + d_2) * \pi_2^* + e^{-\sigma_\varepsilon^2/2} w_1 * \lambda_1$$

We followed the approach of *Olmstead et al.* [2007] and calculated the expected value and elasticity for the other functional forms. For the semi-log function, we obtained:

$$\frac{\partial E(W)}{\partial \theta}\frac{1}{E(W)} = \left(\begin{array}{c} \alpha\left(w_1^* p_1\, \psi_1 + w_2^* p_2\, \psi_2 + w_1(p_1\chi_1 - p_2\chi_2)\right) \\ +\gamma\left(w_2^* d_2\left(\psi_2 - \left(\frac{w_1}{w_2^*}\right)\chi_2\right)\right) \end{array} \right) \Big/ \Omega \qquad (11)$$

Similar procedures can be applied to the linear and SG functions. The only difference is that the error terms are additive. For the linear function the conditional demand is:

$$w_k^*(p_k, y + d_k) = Z\delta + \alpha p_k + \gamma(y + d_k) + \eta + \varepsilon$$

And for the SG:

$$w_k^*(p_k, y + d_k) = Z\delta' + \alpha\left(\frac{(y + d_k)}{p_k}\right) + \gamma\left(\frac{1}{p_k}\right) + \eta + \varepsilon$$

The expected values are [*Moffitt*, 1989]:

$$E(W) = w_1^*\pi_1 + w_2^*\pi_2 + w_1\lambda_1 + \sigma_\eta\left(\phi\left(\frac{w_1 - w_2^*}{\sigma_\eta}\right) - \phi\left(\frac{w_1 - w_1^*}{\sigma_\eta}\right)\right) \qquad (12)$$

Where,

$$\pi_1 = \Phi\left(\frac{w_1 - w_1^*}{\sigma_\eta}\right)$$

$$\pi_2 = 1 - \Phi\left(\frac{w_1 - w_2^*}{\sigma_\eta}\right)$$

$$\lambda_1 = \Phi\left(\frac{w_1 - w_2^*}{\sigma_\eta}\right) - \Phi\left(\frac{w_1 - w_1^*}{\sigma_\eta}\right)$$

And the elasticity is:

$$\frac{\partial E(W)}{\partial \theta}\frac{1}{E(W)} = ((\alpha(p_1\pi_1 + p_2\pi_2) + \gamma\, d_2\pi_2))/E(W) \tag{13}$$

Finally, the elasticity for the SG is:

$$\frac{\partial E(W)}{\partial \theta}\frac{1}{E(W)} = \left(\begin{array}{c} \alpha\left(\frac{1}{p_2}d_2\,\pi_2 \; - \frac{y +\ d_1}{p_1}\ \pi_1 \; - \frac{y\ +\ d_2}{p_2}\ \pi_2\right) \\ -\gamma\left(\frac{1}{p_1}\pi_1\ + \frac{1}{p_2}\ \pi_2\right) \end{array} \right) \Big/ E(W) \tag{14}$$

3.4 Selection criteria

We can use goodness-of-fit criteria to compare functional forms if we are interested in the capacity to explain the variance of the dependent variable. Alternatively, we can use prediction criteria if we are interested in predicting the value of the dependent variable under different scenarios for the explanatory variables. The goodness-of-fit criteria that we calculated are the Akaike information criteria (AIC), defined as:

$$AIC = -2LogL + 2K \tag{15}$$

With $logL$ the logarithmic value of the likelihood function, k, corresponds to the number of parameters in the model. The choice criteria dictate that we choose the model with the lowest AIC. On the other hand, we will use the Mean Square Error as a prediction criterion, which corresponds to the mean value of the squared difference between the predicted value and the observed value of the dependent variable. We estimate the model using 80% of the sample, chosen randomly, and predicted the expected value for the remaining 20% of the sample [*Grootendorst*, 1995]. The statistic is:

$$MSE_j = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_{ij} - y_i)^2 \tag{16}$$

The sum is extended to the individuals of the forecast subsample $\hat{y}_{ij}$, which corresponds to the predicted value of the amount of water consumption by individual ith, using the estimator jth; y_i is the observed water consumption. If the MSE is close to zero, there is no prediction error. However, if the MSE takes on values tending toward a positive infinite the predictive ability is very poor.

3.5 Hypothesis Testing

Our main hypothesis is that there are no differences in the price elasticity (or consumption expected value) between different functional forms. Following [*Turner and Rockel*, 1988] the general hypothesis for our analysis can be written as:

$$\theta = G(\beta, \gamma)$$

where β is a vector of parameters $k \times 1$ of a first functional form $Y_1 = f(X, \beta, \varepsilon_1)$, and γ is a vector of $q \times 1$ parameters estimated from a second functional form $Y_2 = f(Z, \gamma, \varepsilon_2)$; ε_1 and ε_2 are serially independent and homoscedastic for different observations, that is, $E(\varepsilon_{1t}\varepsilon_{2t'}) = 0$, for $t \neq t'$, but correlated for the same observation, with $E(\varepsilon_{1t}\varepsilon_{2t}) = \sigma_{12}$. G is a continuously differentiable function of the parameters β and γ. In this case, G represents the test hypothesis that there is no difference between the expected consumption level (price elasticity) of the two functional forms. Considering that the estimators of maximum likelihood of β and γ are consistent [*Amemiya*, 1985] and that $\hat{\theta} = G(\hat{\beta}, \hat{\gamma})$ is a consistent estimator of $\theta = G(\beta, \gamma)$, the variance for this hypothesis is obtained by the delta method as:

$$V(\hat{\theta}) = g' \Omega g$$

where g is a vector $(k + q) \times 1$ of first partial derivatives (or gradient) of G, with respect to β and γ, and Ω is a matrix $(k + q) \times (k + q)$ of asymptotic variances and covariances equal to:

$$\Omega = \begin{bmatrix} A & C' \\ C & B \end{bmatrix}$$

where A is the $k \times k$ matrix of variances and covariances of $\hat{\beta}$, B is the $q \times q$ matrix of variances and covariances of $\hat{\gamma}$, and C is the qxk matrix of covariances between $\hat{\beta}$ and $\hat{\gamma}$, defined as:

$$C = B\frac{\partial l_2}{\partial \gamma}\left(\frac{\partial l_1}{\partial \beta}\right)' A$$

$\partial l_2/\partial\gamma$ is the gradient vector of the likelihood function of the model 2, and $\partial l_1/\partial\beta$ is the gradient vector of model 1.

4 Data

We used a random sample consisting of a panel of 490 households from the city of Manizales, Colombia, covering water consumption between January 2001 and December 2013.

The price system of residential water in Manizales is an increasing two-block tariff. The first consumption block corresponds to the range that goes from 0 to 20 cubic meters. Consumers must pay an overconsumption tariff if they exceed 20 cubic meters. Additionally, we have information about characteristics of each household, such as number of bathrooms, family size, washing machine available at home, type of housing, and climate variables, such as temperature and precipitation. Table 2 shows descriptive statistics.

Table 2: Descriptive Statistics.

Variable	Average	Std. Dev.	Minimum	Maximum
Household Characteristics				
House (1 for house, 0 otherwise)	0.895	0.307	0	1
Washing Machine	0.873	0.333	0	1
Number of Bathrooms	1.373	0.575	1	4
Family Size	3.574	1.518	1	10
Consumption, price, and income variables				
Consumption	17.566	11.201	1	231
p_1	1132.557	210.622	700.310	1322.560
p_2	1137.548	202.388	850.040	1322.560
w_1	20	0	20	20
$y + d_1$	1260350	868096	587897.3	5896322
$y + d_2$	1260450	868090.8	587897.3	5896322
Climate variables				

Temperature	17.082	0.690	15.250	20.050
Precipitation	181.862	95.799	8.740	541.440

5 Results

As shown in Table 3, our results are in line with the existing literature for the four models tested. All the coefficients are significant at 99% confidence level. We found a positive relationship between water consumption and family size, number of bathrooms, house (versus apartment), and the existence of a washing machine. The parameter of climate variables shows a positive relationship between both temperature/precipitation and water demand.

The Akaike information criterion suggests that the log-log functional form and semi-log functional form have the best goodness of fit, with only a minor difference between them. On the other hand, the MSE suggests that both the linear and the SG models are better for prediction.

Table 3: Results of the discrete-continuous model estimation

	Log-log	Semi-log	Linear	Stone–Geary
Constant	5.807^{***}	3.385^{***}	3.530^{***}	1.796^{***}
	(0.138)	(0.081)	(0.132)	(0.125)
House	0.306^{***}	0.295^{***}	0.313^{***}	0.313^{***}
	(0.009)	(0.009)	(0.015)	(0.015)
Number of Bathrooms	0.076^{***}	0.082^{***}	0.165^{***}	0.166^{***}
	(0.005)	(0.005)	(0.008)	(0.008)
Family Size	0.051^{***}	0.052^{***}	0.058^{***}	0.058^{***}
	(0.002)	(0.002)	(0.003)	(0.003)
Washing Machine	0.119^{***}	0.126^{***}	0.074^{***}	0.077^{***}
	(0.008)	(0.008)	(0.013)	(0.013)
Temperature	-0.048^{***}	-0.050^{***}	-0.091^{***}	-0.092^{***}
	(0.004)	(0.004)	(0.007)	(0.007)
Precipitation	-0.003^{***}	-0.003^{***}	-0.005^{***}	-0.005^{***}

	$(3*10^{-4})$	$(3*10^{-4})$	$(4*10^{-4})$	(0.001)
Price	-0.496^{***}	$-4*10^{-4***}$	-0.001^{***}	
	(0.014)	$(1*10^{-5})$	$(2*10^{-5})$	
Income	0.078^{***}	$3*10^{-5***}$	$3*10^{-6***}$	
	(0.005)	$(3*10^{-6})$	$(5*10^{-7})$	
Income/Price				$3*10^{-5***}$
				$(5*10^{-6})$
1/Price				8.742^{***}
				(0.258)
σ_η	0.478^{***}	0.504^{***}	0.946^{***}	0.884^{***}
	(0.113)	(0.050)	(0.031)	(0.045)
σ_ε	0.462^{***}	0.435^{***}	0.544^{***}	0.639^{***}
	(0.116)	(0.057)	(0.053)	(0.062)
N	63724	63724	63724	63724
AIC	128697.253	128792.105	191662.473	191724.893
MSE	119.523	119.716	118.611	118.661

Table 4 shows the average values of the expected consumption and the price elasticity for each functional form, the standard error calculated through the delta method, and a confidence interval at a 95%.

The average expected value of the water consumption ranges between 17.56 and 18.16 cubic meters, which is a value close to the observed average (17.5 cubic meters). The average value of the elasticity lies between -0.47, for the SG functional form, and -0.56, under the linear functional form. All elasticities are lower than 1 (in absolute value), and, consequently, the price elasticity is inelastic under all functional forms.

Table 4: Average values of the expected consumption and the price elasticity.

	Logarithmic	Semi logarithmic	Linear	Stone–Geary
Expected Value	18.160^{***}	18.164^{***}	17.566^{***}	17.568^{***}
(St. Err.)	(0.054)	(0.054)	(0.043)	(0.043)
C.I. by 95%	$(18.055;\ 18.265)$	$(18.059;\ 18.269)$	$(17.481;\ 17.650)$	$(17.483;\ 17.653)$
Elasticity	-0.495^{***}	-0.531^{***}	-0.561^{***}	-0.477^{***}
(St. Err.)	(0.014)	(0.015)	(0.015)	(0.013)

C.I. by 95%	(−0.522; −0.467)	(−0.561; −0.502)	(−0.591; −0.531)	(−0.502; −0.451)

Standard errors between parentheses. *** $p<0.001$ ** $p<0.01$ * $p<0.05$

Table 5 shows the test statistics for the difference (pair comparisons) of expected consumption among all the functional forms. It is not possible to reject the hypothesis for equality of expected consumption either between the logarithmic and semi-log forms or between the linear and SG forms under 95% confidence level. However, it is possible to state that, between the log-log and the linear functional forms, the expected water consumption average is statistically different, and, that this is also the case between the log–log and SG forms, between the semi-log and linear forms, and between the semi-log and SG forms.

Table 5: Results of the mean difference test for expected consumption.

	Logarithmic	Semi logarithmic	Linear	Stone–Geary
Logarithmic	-	−0.003	0.594***	0.591***
		(−0.05)	(8.65)	(8.61)
Semi logarithmic	-	-	0.597***	0.595***
			(8.67)	(8.65)
Linear	-	-	-	−0.002
				(−0.04)

t-statistics between parenthesis. *** $p<0.001$ ** $p<0.01$ * $p<0.05$

Finally, Table 6 shows the results of the hypothesis test of difference in price elasticity among models. In this case, we reject the hypothesis of equal price elasticity between the log-log and the linear and the SG, between the semi-log and the SG, and between the linear and the SG. We cannot reject the hypothesis of equal price elasticity either between the log–log and semi-log and SG, or between the semi-log and the linear.

Table 6: Results of the mean difference test for the price elasticity.

	Logarithmic	Semi logarithmic	Linear	Stone–Geary
Logarithmic	-	0.036	0.066***	−0.017
		(1.79)	(3.22)	(−0.94)
Semi logarithmic	-	-	0.029	−0.054***
			(1.40)	(2.75)

Linear	-	-	-	-0.084^{***}
				(-4.34)

t-statistics between parenthesis. *** $p<0.001$ ** $p<0.01$ * $p<0.05$

Our estimates differ from the one reported by *Hewitt and Hanemann* [1995], who found values above 1 (in absolute value). Nevertheless, our results are similar to the elasticities reported by *Olmstead et al.* [2007], who found values between -0.609 and -0.331. Compared with the values reported by *Dalhuisen et al.* [2003], our results are in line with more than 80% of the previous literature and also some recent papers, such as *Grafton et al.* [2011], *Polycarpou and Zachariadis* [2012], *Clavijo* [2013], and *Porcher* [2013]. Nevertheless, there are also recent studies that provide more elastic demand functions—see *Miyawaki et al.* [2010], *Miyawaki et al.* [2011] and *Miyawaki et al.* [2014].

From a policy perspective, our hypothesis tests shed light on the relevance of the chosen function form, for both the expected consumption and the price elasticity. In the case of price elasticity, which is a key parameter to assess the welfare effects of water policies, the use of the log-log, the semi-log, and the SG should provide the same information (statistically). The same holds for the selection between the semi-log and the linear form. However, the hypothesis test shows that the use of the log-log and the linear functional forms, in comparison with other functional forms (log–log – linear, log–log – SG, semi-log – SG, and linear – SG) will provide results that are statistically different. This issue is important, considering that these two functional forms are the most used in the literature (see Table 1).

6 Conclusions

We provide evidence that the selection of a functional form for the water demand equation in a discrete-continuous choice model affects the value of both the expected consumption and the price elasticity. We provide evidence using the most common functional forms reported in the literature (linear, semi-log, and log–log) and include a less familiar functional form (SG).

Our results are consistent with most of the previous literature; the expected consumption for all functional forms is around the observed consumption, and the price elasticities are less than 1 (in absolute value), which indicates that water is an inelastic good.

Furthermore, the Akaike information criterion suggests that the log–log and the semi-log functional forms have the best goodness of fit. Nevertheless, the linear and SG functional forms show the best prediction power, measured by mean square error. Therefore, the selection of the appropriate functional form depends on the researcher´s objectives.

Finally, based on the hypothesis test conducted, the selection of the functional form will have consequences for the estimation of key parameters of water demand. To provide better information to both the policy makers and the water utilities companies, we recommend estimating several functional forms reporting a range of values for both the expected consumption and the price elasticities.

Acknowledgements

The authors want to acknowledge the International Development Research Centre (IDRC-Canada), for its financial support through the project "Welfare and Economic Evaluation of Climatic Change Impacts on Water Resources at River Basin Scale: EEC2 – Water Project". (No. 106924-001).

References

Al-Qunaibet, M. H., and R. S. Johnston (1985), Municipal demand for water in Kuwait: methodological issues and empirical results, *Water Resources Research*, *21*(4), 433-438.

Amemiya, T. (1985), *Advanced econometrics*, Harvard university press.

Arbués, F., M. Á. García-Valiñas, and R. Martñinez-Espiñeira (2003), Estimation of residential water demand: a state-of-the-art review, *The Journal of Socio-Economics*, *32*(1), 81-102.

Burtless, G., and J. A. Hausman (1978), The effect of taxation on labor supply: Evaluating the Gary negative income tax experiment, *The Journal of Political Economy*, 1103-1130.

Clavijo, A. (2013), Estimación de la función de demanda por agua potable. Aplicación para la cuenca de Jequetepeque en Perú, Universidad de Chile.

Dalhuisen, J. M., R. J. Florax, H. L. De Groot, and P. Nijkamp (2003), Price and income elasticities of residential water demand: a meta-analysis, *Land economics*, *79*(2), 292-308.

Ferrara, I. (2008), Residential water use, *OECD Journal: General Papers*, *2008*(2), 153-180.

Fuente, D., J. Gakii Gatua, M. Ikiara, J. Kabubo-Mariara, M. Mwaura, and D. Whittington (2016), Water and sanitation service delivery, pricing, and the poor: An empirical estimate of subsidy incidence in Nairobi, Kenya, *Water Resources Research*.

Gaudin, S., R. C. Griffin, and R. C. Sickles (2001), Demand specification for municipal water management: evaluation of the Stone-Geary form, *Land Economics*, *77*(3), 399-422.

Grafton, R. Q., M. B. Ward, H. To, and T. Kompas (2011), Determinants of residential water consumption: Evidence and analysis from a 10-country household survey, *Water Resources Research*, *47*(8), n/a-n/a.

Grootendorst, P. V. (1995), A comparison of alternative models of prescription drug utilization, *Health Economics*, *4*(3), 183-198.

Headley, J. C. (1963), The relation of family income and use of water for residential and commercial purposes in the san francisco-oakland metropolitan area, *Land Economics*, 441-449.

Hewitt, J. A., and W. M. Hanemann (1995), A discrete/continuous choice approach to residential water demand under block rate pricing, *Land Economics*, 173-192.

Howe, C. W., and F. P. Linaweaver (1967), The impact of price on residential water demand and its relation to system design and price structure, *Water Resources Research*, *3*(1), 13-32.

Miyawaki, K., Y. Omori, and A. Hibiki (2010), Bayesian estimation of demand functions under block rate pricing, *CIRJE Discussion Papers*, *January*.

Miyawaki, K., Y. Omori, and A. Hibiki (2011), Panel Data Analysis of Japanese Residential Water Demand Using a Discrete/Continuous Choice Approach, *Japanese Economic Review*, *62*(3), 365-386.

Miyawaki, K., Y. Omori, and A. Hibiki (2014), Exact Estimation of Demand Functions under Block-Rate Pricing, *Econometric Reviews*(ahead-of-print), 1-33.

Moffitt, R. (1986), The econometrics of piecewise-linear budget constraints: a survey and exposition of the maximum likelihood method, *Journal of Business & Economic Statistics*, *4*(3), 317-328.

Moffitt, R. (1989), Estimating the value of an in-kind transfer: The case of food stamps, *Econometrica: Journal of the Econometric Society*, 385-409.

Nordin, J. A. (1976), A proposed modification of Taylor's demand analysis: comment, *Bell Journal of Economics*, *7*(2), 719-721.

Olmstead, S. M., W. M. Hanemann, and R. N. Stavins (2007), Water demand under alternative price structures, *Journal of Environmental Economics and Management*, *54*(2), 181-198.

Polycarpou, A., and T. Zachariadis (2012), An Econometric Analysis of Residential Water Demand in Cyprus, *Water Resources Management*, *27*(1), 309-317.

Porcher, S. (2013), Efficiency and equity in two-part tariffs: the case of residential water rates, *Applied Economics*, *46*(5), 539-555.

Taylor, L. D. (1975), The demand for electricity: a survey, *The Bell Journal of Economics*, 74-110.

Turner, R. W., and M. L. Rockel (1988), Estimating covariances of parameter estimates from different models, *Economics Letters*, *26*(2), 137-140.

Wong, S. (1972), A model on municipal water demand: a case study of Northeastern Illinois, *Land Economics*, 34-44.

Young, R. A. (1973), Price elasticity of demand for municipal water: a case study of Tucson, Arizona, *Water Resources Research*, *9*(4), 1068-1072.