CONTRIBUTI DI RICERCA CRENOS

CRENoS
CENTRO RICERCHE
ECONOMICHE NORD SUD
**Università di Cagliari**
**Università di Sassari**

# APPLICATION OF THE ML HAUSMAN APPROACH TO THE DEMAND OF WATER FOR RESIDENTIAL USE: HETEROGENEITY vs TWO-ERROR SPECIFICATION

Elisabetta Strazzera

WORKING PAPERS

CUEC

2006/04

**CENTRO RICERCHE ECONOMICHE NORD SUD**
**(CRENOS)**
**UNIVERSITÀ DI CAGLIARI**
**UNIVERSITÀ DI SASSARI**

Il CRENoS è un centro di ricerca istituito nel 1993 che fa capo alle Università di Cagliari e Sassari ed è attualmente diretto da Raffaele Paci. Il CRENoS si propone di contribuire a migliorare le conoscenze sul divario economico tra aree integrate e di fornire utili indicazioni di intervento. Particolare attenzione è dedicata al ruolo svolto dalle istituzioni, dal progresso tecnologico e dalla diffusione dell'innovazione nel processo di convergenza o divergenza tra aree economiche. Il CRENoS si propone inoltre di studiare la compatibilità fra tali processi e la salvaguardia delle risorse ambientali, sia globali sia locali.
Per svolgere la sua attività di ricerca, il CRENoS collabora con centri di ricerca e università nazionali ed internazionali; è attivo nell'organizzare conferenze ad alto contenuto scientifico, seminari e altre attività di natura formativa; tiene aggiornate una serie di banche dati e ha una sua collana di pubblicazioni.

**www.crenos.it**
**info@crenos.it**

CRENOS – CAGLIARI
VIA SAN GIORGIO 12, I-09100 CAGLIARI, ITALIA
TEL. +39-070-6756414; FAX +39-070- 6756402 / 6753760

CRENOS - SASSARI
VIA TORRE TONDA 34, I-07100 SASSARI, ITALIA
TEL. +39-079-2017301; FAX +39-079-2017312

Titolo: APPLICATION OF THE ML HAUSMAN APPROACH TO THE DEMAND OF WATER FOR RESIDENTIAL USE: HETEROGENEITY vs TWO-ERROR SPECIFICATION

ISBN 88 – 8467 – 328 - 3

Prima edizione Aprile 2006

Via Is Mirrionis, 1
09123 Cagliari
Tel./Fax 070291201
www.cuec.it

# Application of the ML Hausman approach to the demand of water for residential use: heterogeneity vs two-error specification

**Elisabetta Strazzera**[*]
*DRES and CRENoS, University of Cagliari*

## Abstract

This paper presents an application of two ML models to the analysis of residential demand of water: the heterogeneity and the two-error model, both apt to model demand in presence of a kinked budget constraint. The heterogeneity model is especially suitable when the distribution is characterized by a strong clustering around the kinks. Since in practice observations can be very close, but not exactly at the kink, its application may require the definition of an interval of data around the kink, so that the observations falling inside this interval are attributed to the kink. We propose a procedure, based upon the estimates obtained from the two-error model, to define this interval. In this application we find that the heterogeneity model allows to obtain more efficient estimates than the two-error model for the parameter of principal interest, i.e. the coefficient of the price variable.

**Keywords: Water Demand, Block Pricing, Kinked Budget Constraint, Maximum Likelihood, Discrete-Continuous Choice, Hausman model**

JEL classification: C24, C51, D12, Q25.

---

[*] This research was set up while visiting the University of California, Berkeley. I would like to thank, without implicating, Michael Hanemann for very kind hospitality at the Agricultural and Resource Economics department. The data used in this paper were collected as a part of a study funded by the Ministry of Education of Italy, PON RIADE project, principal investigator ENEA.

## 1. Introduction

A substantial stream of research has been devoted in the last decade to the problem of analyzing the functional form of the demand of water, in order to assess the effect of changes in the supply function. Residential consumption competes for this increasingly scarce resource with other sectors: agriculture, industry, tourism. An efficient supply allocation would require identification of the determinants of the residential demand in order to control non necessary uses, i.e. all uses exceeding the normal requirements for alimentary and hygienic needs. Market instruments may be helpful to control such unessential uses of water, and rising blocks schemes are advocated, and currently widely adopted, as an effective strategy to accomplish this aim.

The problem is that the econometric analysis of the demand function in presence of a block price is quite complex. The fundamental work of Moffitt (1986,1990) has clearly exposited the correct strategy to model demand in presence of a piecewise linear budget constraint: demand is conditional on the choice of a specific portion of the budget constraint, and the two choices (continuous-discrete) should be modeled as a joint process. Based upon a seminal paper by Burtless and Hausman (1978), Moffitt proposes two alternative Maximum Likelihood models to account for non linearity and endogeneity in the price structure for this type of setting. Unfortunately, due to non convexities of the likelihood function, maximization is often difficult: the procedures often break down, multiple local optima can be found, and this requires using several starting point vectors to find the global maximum. This has induced many researchers to apply alternative models (Instrumental Variables, Two-part models), which, although not optimal, can be more easily estimated. As Nauges and Thomas (2000) remark: "results on important figures such as price and income elasticities are very heterogeneous and remain often sensitive to econometric specification". Finding a robust method to ensure reliable estimates is a crucial issue for the current research in this field.

This paper presents the application of two Hausman models –as exposed by Moffitt (1986)- to the analysis of residential demand of water: the heterogeneity and the two-error model. The first application in this context is due to Hewitt (1993) and Hewitt and Hanemann (1995). While the two error model has been used in other work on residential demand of water (Pint, 1999, Rietveld et al., 1997, Olmstead et al., 2005), and on the demand of water for irrigation (Bar-Shira et al., 2005), we are not aware of any application of the heterogeneity model to water demand subsequent to Hewitt (1993). It will be seen that this model is especially apt to fit distributions where a strong clustering around the kinks is observed. In practice, many observations might be close to the kink but not *exactly* at the kink: in order to apply this model, an interval of data around the kink should be defined (see Moffitt and Nicholson, 1982; Friedberg, 2000), and the observations falling inside this interval should be attributed to the kink. We propose a procedure, based upon the estimates obtained from the two-error model, to define this interval. It will be seen that in this application the heterogeneity model allows to obtain more efficient estimates than the two-error model for the parameter of main interest, i.e. the coefficient of the price variable.

## 2. Economics of the Kinked Budget Constraint

The consumer's problem can be modeled as follows: the utility function U(·) is a function of the consumed quantity of the good of interest (in our case, drinking water), defined as $q_1$, and all other goods x. The good q is assumed to be a normal good. A block structure may be increasing, if price blocks are increasing in quantity, or, alternatively, decreasing. Since the general trend for utilities like water and energy is to adopt increasing block tariff structures, which induce more conservative usage of the resource, we restrict attention to this case only. Let us consider for simplicity a two block structure, as follows:

$$P(q) = p1 \ (q \leq k)$$
$$P(q) = p2 \ (q > k)$$

where k is the value of q at the kink (block limit); p1 is the first block price, p2 is the second block price, and p2 > p1 (increasing blocks).
Let x be a composite good in the choice set alternative to $q_1$, with price normalized at 1. The consumer maximizes U(·) subject to the following budget constraint, which is non linear in q1:

$$min(k,q)\, p1 + max(q-k,0)p2 + x \leq y \qquad (1)$$

where y is the consumer's income. Eq. (1) can also be written as follows:

$$q\,p1 + x \le y \;(q \le k) \qquad (2)$$
$$qp2 + x \le y + k(p2 - p1) = yv \;(q > k)$$

where yv is defined as the consumer's *virtual income* in the region {q>k}, i.e. income that would be just about sufficient to buy the bundle q, x at the prices p2 and 1. The virtual income is represented in the following graph as the intercept of the dashed budget line on the price-axis.

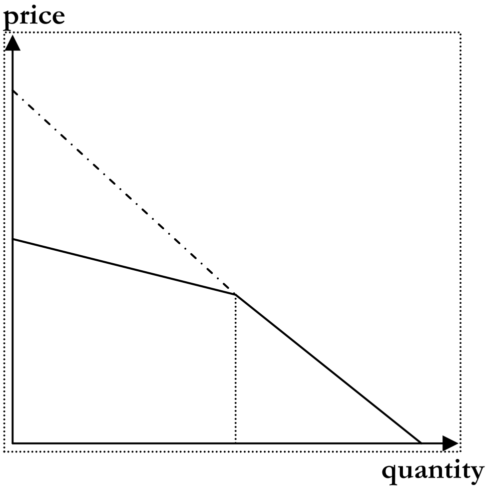

**Fig. 1.**

Let's define the consumer's indirect utility function:

$$V=V(y,P(q)) = U(g(P(q),y)\,,\,y-g(P(q),y)),$$

where g(·) is the standard Hicksian demand function, characterized by the following structure:

$$g(p1,y) \text{ if } g(p1,y) \le k \text{ and } V(y,p1) > V(yv,p2) \qquad (3)$$
$$g(p2,yv) \text{ if } g(p2,yv) > k \text{ and } V(yv,p2) > V(y,p1)$$
$$k \text{ otherwise.}$$

As Moffitt (1986) shows, (3) easily transforms into the following:

$$q = d1\, g(p1,y) + d2\, g(p2,yv) + (1-d1-d2)\, k \tag{4}$$

$d1$ is 1 if $k > g(p1,y)$, 0 otherwise;

$d2$ is 1 if $k < g(p2,yv)$, 0 otherwise.

The demand function (4) is represented in the following two graphs, where different possible equilibria are shown.

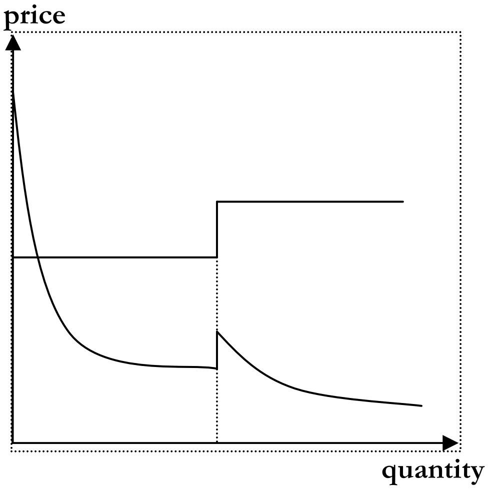

**Fig. 2**

Figure 2 represents a “regular” solution, where the intersection between the individual demand function and the price occurs along the first segment of the price line. In such a situation (and analogously if this had happened in the second segment of the price line) there is a unique equilibrium, which is determined by the textbook equivalence condition of marginal rate of substitution equal to price ratio.

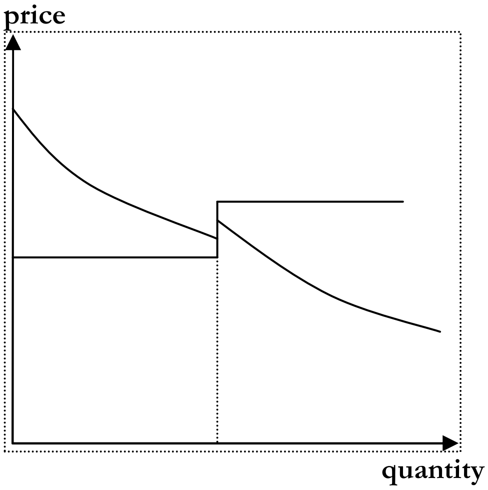

**Fig. 3.**

In contrast, figure 3 represents a situation where the demand function crosses the price curve at the limit of the block price, i.e. at the kink point of the budget line in Figure 1. At this point the demand function is not defined, and different marginal rates of substitution may be compatible with the solution at the kink point. This means that different values of income, prices, and other variables argument of the demand function are compatible with the same level of water demanded. For example, in Figure 1 the same quantity at the kink is compatible with both the actual and virtual individual's income. This has relevant implications on the distribution of the demand of water, as the kink point results in a mass point, and this has to be properly taken into account in the econometric modeling.

## 3. Econometrics of the Kinked Budget Constraint

In the previous section we have seen that when tariffs have a block structure, demand is a non linear function of income, prices and the other relevant variables argument of the demand function. Moreover, it is possible that data are clustered round the kinks, since at a kink point different marginal rates of substitution, and hence different levels of the variables argument of the demand function, are compatible with that level of demand.

A simple econometric specification for the demand equation is the following:

$$q = \beta X + \gamma pm + \delta yv + \varepsilon \qquad (5)$$

where X is a vector of explanatory variables other than income and prices; $\beta$, $\gamma$ e $\delta$ are parameters to be estimated, and $\varepsilon$ is the error term. The equation is not linear in pm nor in yv , since both are a function of the actual level of consumption. Overlooking the non linearity problem, and applying the OLS method to estimate the demand equation, leads to biased and inconsistent estimates, since pm e yv are correlated with the error term $\varepsilon$ (as they depend on q). Moreover, in case of increasing blocks price structure, the OLS method gives an estimate of the price coefficient which is positive, as price is, by construction, positively correlated to the demanded quantity. This specific problem may be solved applying Instrumental Variables (IV) methods, or Heckman's two-stage sample selection method. However, as discussed in Moffitt (1991), these models do not address in a satisfactory way the problem of mass point (clusters at kink) estimation. In their seminal work, Burtless and Hausman (1978) apply the Maximum Likelihood (ML) approach to a selectivity model to take account of both non linearity and discontinuity problems.

Moffitt (1986) examines two alternative "Hausman" ML models, and the conditions under which either is best suited to fit the data. When the data show a strong clustering around the kink points it may be preferable to use a model that assumes that individuals are maximizing there, assigning observations unambiguously to the kink points or the segments where they are found. In this case different locations along each segment are explained by some unobservable factor (i.e., unobservable to the econometrician, not to the consumer), or, in other words, to some *heterogeneity* of preferences. Conversely, when the empirical distribution does not show a strong clustering of data around the kink points, Moffitt suggests the use of a model that takes into account the possibility of a discrepancy between the point where a consumer intends to maximize, and the point where the individual observation is actually found. This may be due to different sources of errors: one possibility is that measurement of consumption is not accurate (measurement error); or some factor exogenous to the consumer's decision making (as modeled by the demand function) determines a different level of consumption from what planned (optimization error); or, last but not least, to some specification error made by the econometrician.

Based upon equation (4), the heterogeneous (one error) econometric model has the following probabilistic structure:

$$\Pr(q = g(p_1, y) + \eta) \qquad \textit{for observations in segment 1}$$
$$\Pr(q = g(p_2, y_v) + \eta) \qquad \textit{for observations in segment 2}$$
$$\Pr(q = g(p_1, y) + \eta) < q^* < \Pr(q = g(p_2, y_v) + \eta)$$
$$\textit{for observations at the kink}$$

A characteristic feature of this model is that the probability of consumption at a certain level is computed for each observation conditioning on the observed location on a segment or kink: in other words, there is *sample separation*. Alternatively, the two error model derives the *unconditional* probability for each individual to attain a certain level of consumption, independently of the level of observed consumption. Its probabilistic structure is the following:

$$\begin{aligned}&\Pr\{[q = g(p_1, y) + \eta + \varepsilon], [g(p_1, y) + \eta < q^*]\} + \\ &\Pr\{[q = g(p_2, y_v) + \eta + \varepsilon], [g(p_2, y_v) + \eta > q^*)]\} + \\ &\Pr\{[q = q^* + \varepsilon], [g(p_1, y) + \eta < q^* < g(p_2, y_v) + \eta]\}\end{aligned} \tag{6}$$

Each term in the addition is the joint probability that an observation is found in a specific section of the budget (first segment, second segment, or kink) *and* at a certain level of consumption inside that section. The probabilities are calculated for every observation in the sample, i.e. there is no sample separation in this case.

## 4. Model specification

Estimation of the two models (5) and (6) requires specification in both the structural part, i.e. the demand function, and the probability distribution of the error terms. As in Hewitt and Hanemann (1995), we assume that the consumer maximizes an indirect utility function defined as:

$$V(p_1, y) = \exp(-\mu\ p_1)\left[ y + \frac{1}{\mu}\left( \gamma p_1 + \frac{\gamma}{\mu} + z'\delta \right) \right]$$

which implies the demand function:

$$g(p_1, y) = z'\delta + \gamma p_1 + \mu y .$$

The latter equation differs with respect to (4) in that it contains the term $z'\delta$, which refers to socio-economic variables that may be included in the model. As discussed in Hewitt (1993, p.79), this implies that the socio-economic variables enter as an argument of the utility function: whether this is a reasonable assumption depends on the specific set of covariates used.

The heterogeneity model is obtained by applying an error term to the conditional demands –i.e. conditional to the observed location on the budget line:

$$g^k(p_1, y) = z'\delta + \gamma p_{1k} + \mu(y^k) + \varepsilon$$

where *k* is the block where demand is observed, and income is real income for individuals in the first block, and virtual income for individuals in higher blocks of consumption.

Considering now a two blocks tariff structure, which is what we will deal with in our application, the likelihood function is:

$$\ell_h = \sum_{Block\,1} [\phi(w_1)] + \sum_{Kink} \ln[\Phi(t_2) - \Phi(t_1)] + \sum_{Block\,2} [\phi(w_2)]$$

where $\phi$ and $\Phi$ are, respectively, the Normal density and the distribution function, which we assume for the error term, with mean zero and variance $\sigma^2$. Although other distributions may be assumed, the Normal generally represents a quite convenient choice; a typical expedient used to deal with asymmetric distributions is to assume a Log-normal distribution for the error term, and work with the logarithms of the dependent variable, which are distributed as a Normal. The other terms are:

$$w_k = \frac{\left[q - z'\delta - \mu(y^k) - \gamma p_k\right]}{\sigma}$$

$$t_k = \frac{\left[q^* - z'\delta - \mu(y^k) - \gamma p_k\right]}{\sigma}$$

Here the data observed in each section (segment or kink) of the budget line enter the likelihood separately from the others.

The two error model is more complex, since it requires computing the joint probability for each observation to occur in a particular segment or kink, and to realize a specific level of consumption.

Given the two error stochastic structure, we need to specify a bivariate distribution for the joint probabilities of $\eta$ and $v=\eta+\varepsilon$: the Bivariate Normal is a natural candidate, and the corresponding log-likelihood function derived by Moffitt (1986) is:

$$\ell_{2err} = \sum_{All}\left[\frac{1}{\sigma_v}\phi(w_1)\Phi(r_1) + \frac{1}{\sigma_\eta}\phi(v)\left[\Phi(t_2) - \Phi(t_1)\right] + \frac{1}{\sigma_v}\phi(w_2)\left[1 - \Phi(r_2)\right]\right]$$

where

$$w_k = \frac{\left[q - z'\delta - \mu(y_v) - \gamma p_k\right]}{\sigma_v},$$

$$t_k = \frac{\left[q^* - z'\delta - \mu(y_v) - \gamma p_k\right]}{\sigma_\varepsilon},$$

$$r_k = \frac{t_k - \rho w_k}{\sqrt{1-\rho^2}}, \qquad \nu = \frac{[q - q^*]}{\sigma_\eta},$$

$$\sigma_\nu = \sqrt{\sigma_\varepsilon^2 + \sigma_\eta^2}, \qquad \rho = \frac{\sigma_\varepsilon}{\sigma_\nu}.$$

In this model there is no sample separation: for each observation there exists some probability that it lies in some other section of the budget line than that where it was actually observed.

## 5. Structural determinants of the demand of water

As said in the previous section, the regression usually contains some socioeconomic or “structural” components –other than income and prices- entering the demand function of water for residential uses. Most obviously the household size affects the level of consumption, and incomplete information on this feature would result in a serious omitted variable problem, and larger unexplained heterogeneity. Many studies use structural characteristics of the house, mostly intended as proxies for the household size: number of rooms, number of bathrooms, or the dimension of the building. Use of appliances, like laundry machines, dishwashers, hot tubs, may in some contexts be relevant to explain different levels of consumption. Garden size, the system of irrigation and its frequency, climate features (in panel data, or when cross sections comprise various climatic areas) are also included as regressors in many water demand models: the reader is referred to Arbués et al. (2003) for an overview of the covariate specifications found in the literature.

## 6. Application of the Hausman ML models to a cross section analysis of water demand

In this work we will see an application of the Hausman model to data relative to water consumption in Alghero, a marine town in Sardinia, Italy, which is characterized by a heavy tourist load during the summer months (the population density in the July-August months increases by

150%) and an excess demand of water for domestic use, which is currently controlled through quantity restrictions.

A survey was taken over a sample of 404 households, drawn from a population of residential customers endowed with a single unit metering device. The individual consumption data was provided by the municipal water company, while the survey was intended to collect information on household socioeconomic and demographic characteristics, and on structural features of the dwelling. Table 1 in Appendix reports the summary statistics of the variables used in the analysis.

The price of water is structured as an increasing block tariff, with a fixed access charge, a first block price of €0.62 for annual consumption up to 160 m$^3$, and €0.92 for higher levels of consumption.

The following figure represents the histogram of the annual water consumption per family.

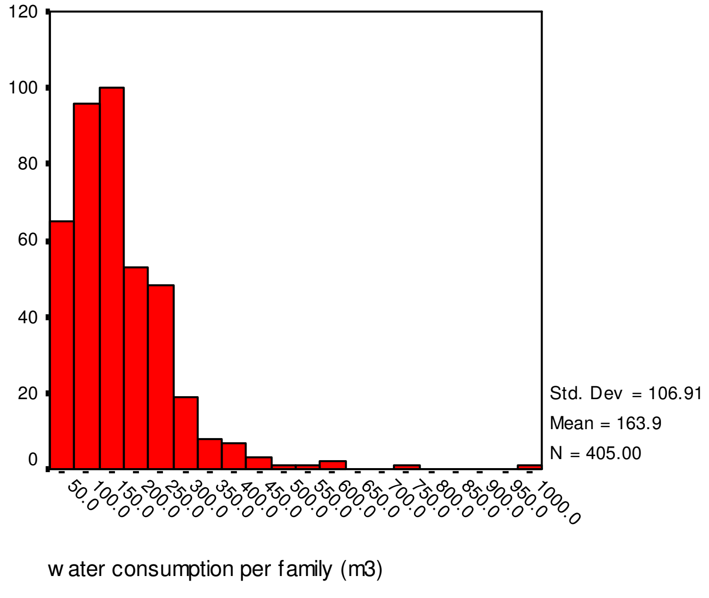

**Fig.4**

For the analysis of these data we apply the two Hausman models exposited by Moffitt (1986): the heterogeneity univariate model; and the bivariate model, that allows estimation of both heterogeneity and measurement errors. As discussed above, the heterogeneity model assumes that each consumer selects the block (or kink) based on a correct maximization process; along each segment, individuals may choose different levels of consumption due to unobservable heterogeneity among individuals. This model implies that each consumer is found exactly in the portion of the budget constraint that is compatible with his/her preferences. Since the kink point is compatible with many different preference structures, this model predicts some clustering of observations around kink points, hence it is especially apt to fit distributions characterized by such clusters. Alternatively, the two-error model allows some discrepancy between planned and real levels of consumption. Such discrepancy may be due to optimization or perception errors (for example, because imperfect information on the tariff structure), or by measurement errors (for example, because some measurement instruments are faulty). If this is the correct scenario, individuals who planned to consume at a certain block may be found to consume at a different block. This model is better suited to situations where observations are spread over the support, and no clusters are observed around the kink points.

Our data show two strong clusters: the first at about 100m$^3$, the second, with a higher peak, at about 150m$^3$. The heterogeneity model could be considered more suitable for this type of data; however, a practical problem arises when it comes to define the "kink" observations. If only the observations lying exactly at the kink, i.e. 160m$^3$, are modeled as kink observations, then in our case only one observation would respond to the requisite. Besides being overly restrictive (it is reasonable to concede some rounding error around the optimization point even if we are not willing to allow for large departures from this point), this strategy leads to serious problems in the estimation of the model, as it will be seen later.

On the other hand, assigning observations to the kink might be seen as an arbitrary manipulation of the data. Friedberg (2000) assigns observations to the kink based on some *ad hoc* judgment; Moffitt and Nicholson (1982) use different bandwidths around the kink to group observations, and then select the model with the most "plausible" epsilon based on some *a priori*, and through likelihood ratio tests. While using a priori information may be helpful in general, we will see that looking at the likelihood values may be misleading if –as happens- the likelihood function is not well behaved.

The asymmetric shape of the water consumption distribution suggests that a Lognormal specification for the error terms would fit the data better than the Normal. The dependent variable is therefore transformed into logarithms, and the same transformation is operated on the continuous independent variables, which are: marginal price; (virtual) income; household size; information; years in the house; house at ground floor or upper floors. The first three variables do not require special comments; "information" is a category variable signaling if the consumer thought he or she was paying less (negative sign) or more (positive sign) than what was effectively paid; higher numbers correspond to larger differences between the stated amount and the real amount. We expect that people who think they pay more than they actually do will be more conservative in the use of water (and vice-versa): this implies that we expect a negative sign for the coefficient of this variable. The variable "years in the house" is a proxy for how old the house and its technological equipment is: we could not use the real variable since many people were not able to respond to this specific question. We expect that the older the house, the higher the consumption (mainly because of old flushing systems, and a higher probability of leakages). Finally, the variable "low ground" includes all houses, and apartments located at low ground, which are often endowed with some open space (backyard or patio) and we expect a positive sign for the coefficient.

**Table 1: Two error and heterogeneity models, original data**

| **Dep=ln(m$^3$)** | **Two error model** | | **Heterogeneity model** | |
|---|---|---|---|---|
| **n.obs=385** | ***Coefficients (std errors)*** | ***P-value*** | ***Coefficients (std errors)*** | ***P-value*** |
| **Constant** | 3.474<br>(0.293) | 0.000 | 5.176<br>(0.130) | 0.000 |
| **M. Price** | -0.538<br>(0.363) | 0.138 | 2.109<br>(0.101) | 0.000 |
| **Income** | 0.149<br>(0.027) | 0.000 | 0.043<br>(0.016) | 0.007 |
| **HH size** | 0.270<br>(0.073) | 0.000 | 0.084<br>(0.042) | 0.045 |
| **Information** | -0.125<br>(0.020) | 0.000 | -0.047<br>(0.010) | 0.000 |
| **House yrs** | 0.094<br>(0.042) | 0.026 | 0.031<br>(0.025) | 0.207 |
| **Low ground** | 0.094<br>(0.075) | 0.214 | 0.076<br>(0.044) | 0.085 |
| $\sigma_\eta$ | 0.501<br>(0.053) | 0.000 | 0.360<br>(0.013) | 0.000 |
| $\sigma_\varepsilon$ | 0.291<br>(0.065) | 0.000 | -- | -- |
| **Mean log-lik** | -0.769 | | -0.545 | |
| **n. log-lik>0** | 0 | | 131 | |

The poor performance of the heterogeneity model is apparent from two features: first, the price coefficient has a positive sign, which means that the model does not succeed in treating the endogeneity of price due to the increasing block structure; second, a relevant number of individual log-likelihoods is positive, which is an evident sign of problems in the likelihood function. In this case, the two-error model does not show any problems: a close inspection of the likelihood function at the maximum shows that all values are well inside the boundaries. This does not imply, though, that the two-error model is in general immune to this type of failure: similar problems for this model have been reported by Herriges and King (1994), and by Rietveld et al. (1997).

The two-error fit seems quite satisfactory, since all coefficients have the expected sign and most of them show p-values close to zero. Both standard errors are significantly different from zero, and this could lead to accept the two-error specification. A sensitivity analysis was carried out to check the robustness of the estimates to small modifications in the sample size and the covariate specification, with positive results. However, the price coefficient, i.e. the most important parameter in the context of the present study, is not estimated efficiently. Over-parameterization is recognized to be a problem for the two-error model (Moffitt, 1986, p. 326; Heckman and Singer, 1984), which may cause a decrease in the precision of some parameter estimates; and this may especially be so when the sample size is small as in this case. As discussed above, a possible strategy is to apply the heterogeneity model to manipulated data, where observations that are close enough to the kink are assigned to it. We used the estimated standard error $\sigma_\varepsilon$ to define a bandwidth around the kink: observations in the interval $\{\ln(160) \pm \alpha \cdot \sigma_\varepsilon\}, \alpha \in (0,1)$, were assigned to the kink, and the heterogeneity model was estimated for each value of the grid. The punctual estimates from the heterogeneity model get closest to those obtained from the two-error model at $\alpha$=0.35, and all coefficients are significant. The number of observations attributed to the kink is 49, lying in the interval [145,177]. We do not report all estimates obtained from the grid, but just observe that the price coefficient monotonically increases with $\alpha$, reaching the value -1.94 for $\alpha$=0.9.

**Table 2: Two error and heterogeneity models, cluster data assigned to the kink**

| **Dep=ln(m$^3$)** | **Two error model** | | **Heterogeneity model** | |
|---|---|---|---|---|
| **n.obs=385** | ***Coefficients (std errors)*** | ***P-value*** | ***Coefficients (std errors)*** | ***P-value*** |
| **Constant** | 3.469<br>(0.291) | 0.000 | 3.456<br>(0.194) | 0.000 |
| **M. Price** | -0.544<br>(0.360) | 0.131 | -0.542<br>(0.080) | 0.000 |
| **Income** | 0.149<br>(0.027) | 0.000 | 0.152<br>(0.025) | 0.000 |
| **HH size** | 0.271<br>(0.074) | 0.000 | 0.272<br>(0.069) | 0.000 |
| **Information** | -0.125<br>(0.020) | 0.000 | -0.121<br>(0.015) | 0.000 |
| **House yrs** | 0.095<br>(0.042) | 0.024 | 0.097<br>(0.041) | 0.017 |
| **Low ground** | 0.094<br>(0.075) | 0.211 | 0.084<br>(0.074) | 0.252 |
| $\sigma_\eta$ | 0.506<br>(0.052) | 0.000 | 0.597<br>(0.024) | 0.000 |
| $\sigma_\varepsilon$ | 0.285<br>(0.067) | 0.000 | | -- |
| **Mean log-lik** | -0.768 | | -1.097 | |
| **n. log-lik>0** | 0 | | 0 | |

Our benchmark was the two-error model estimated from the original data, but it can be observed that it gives very similar results when estimated on the manipulated data. However, for higher values of $\alpha$ this does not hold anymore, and the two-error model shows specification problems (the routine breaks down; the log-likelihood takes positive values for some observations).

## 7. Conclusions

Accurate estimation of the demand of water is essential for a correct planning and management of a resource that becomes more and more scarce. It is widely recognized that the ML Hausman approach is the most appropriate to model a demand function in presence of a kinked budget constraint; however, such approach has been adopted in very few empirical works. The problem is that estimation may be difficult, because the likelihood function is not globally concave: but use of different starting values vectors, inspection of the likelihood function in the neighborhood of the global maximum, and sensitivity analysis can greatly help to select a valid specification. In the present work, the two-error model was applied to data on water demand, and, conditional on the selected specification, the maximization procedure did not present any problems. However, the coefficient of the price variable, which was a key parameter in this study, is not estimated efficiently by this model. As discussed by Moffitt (1986), when the distribution of the dependent variable is characterized by a strong clustering of observations around the kink, the heterogeneous model may be a better candidate than the two-error model to fit such distribution. Since many observations are close to the kink, even though not exactly at the kink, it is necessary to attribute a number of observations to the kink in order to apply the heterogeneity model. We propose a criterion to classify data as kink observations. The heterogeneity model applied to the modified data produces efficient estimates for all parameters of major interest.

## References

Bar-Shira, Z., I. Finkelshtein and A. Simhon (2005), "Regulating irrigation via Block-Rate Pricing: an Econometric Analysis", Discussion paper n.05.01, The Maurice Falk Institute for Economic Research in Israel.

Burtless, G. and J. Hausman (1978), "The Effect of Taxation on Labor Supply: Evaluating the Gary Income Maintenance Experiment", Journal of Political Economy, 86: 1103:1130.

Friedberg, L. (2000), "The Labor Supply Effects of the Social Security Earnings Test", The Review of Economics and Statistics, 82 (1): 48-63.

Herriges, J.A. and K.K. King (1994), "Residential Demand for Electricity under Inverted Block Rates: Evidence from a Controlled Experiment", Journal of Business and Economic Statistics, 12(4): 419-30.

Hewitt, J.A. and W.M. Hanemann (1995), "A Discrete/Continuous Choice Approach to Residential Water Demand under Block Rate Pricing", Land Economics, 71(2): 173-92

Hewitt (1993), "Watering Households: The Two-error Discrete/Continuous Choice Model of Residential Water Demand", PhD Dissertation, University of California, Berkeley.

Moffitt, R. (1986), "The Econometrics of Piecewise-Linear Budget Constraints", Journal of Business and Economic Statistics, 4(3): 317-28.

Moffitt (1990), "The Econometrics of Kinked Budget Constraints", Journal of Economic Perspectives, 4(2): 119-39.

Moffitt R. and W. Nicholson (1982), "The Effect of Unemployment Insurance on Unemployment: the Case of Federal Supplemental Benefits", The Review of Economics and Statistics, 64(1): 1:11.

Nauges, C. and A. Thomas (2000), "Privately Operated Water Utilities, Municipal Price Negotiation, and Estimation of Residential Water Demand: the Case of France", Land Economics, 76(1): 68-85.

Olmstead, S.M., W.M. Hanemann and R.N. Stavin (2005), "Do Consumers React to the Shape of Supply? Water Demand Under Heterogenous Price Structures", Regulatory Policy Program working paper RPP 2005-05, J.F. Kennedy School of Government, Harvard University.

Pint, E.M. (1999), "Household Responses to Increased Water Rates During the California Drought", Land Economics, 75(2): 246-66.

Rietveld, P., J. Rouwendal and B. Zwart (1997), "Estimating Demand in Urban Indonesia: a Maximum Likelihood Approach to Block Rate Pricing Data", mimeo, Vrije Universiteit Amsterdam

**Appendix**

**Table A.1**

| REGRESSORS | Mean (std) | Description |
|---|---|---|
| **HH SIZE** | **1.02<br>(0.46)** | log(n. of people in the house) |
| **HOUSE YEARS** | **2.91<br>(0.76)** | log(years in the house) |
| **INCOME** | **4.99<br>(1.23)** | log(grocery expenditures / n. of people in the house) |
| **INFORMATION** | **0.33<br>(2.02)** | Scalar, range [-4,+4]: difference between real water bill and what respondents think they paid |
| **LOW GROUND** | **0.23<br>(0.42)** | 1: low ground; zero: otherwise |
| **$M^3$** | **4.93<br>(0.60)** | log(cubic meters of water consumed by the unit per yr) |
| **MARGINAL PRICE** | **log (0.62)** | First block (up to 160 mc) |
|  | **log (0.92)** | Second block (over 160 mc) |

## Ultimi Contributi di Ricerca CRENoS

*I Paper sono disponibili in:* http://www.crenos.it

**06/03** *Silvia Loddo,* "Structural Funds and Regional Convergence in Italy"

**06/02** *Fabio Cerina,* "Tourism specialization and environmental sustainability in a dynamic economy"

**06/01** *Emanuela Marrocu, Raffaele Paci e Francesco Pigliaru,* "Gli effetti del capitale pubblico sulla produttività delle regioni italiane"

**05/14** *Rinaldo Brau and Davide Cao*, "Uncovering the macrostructure of tourists' preferences. A choice experiment analysis of tourism demand to Sardinia"

**05/13** *Barbara Dettori, Raffaele Paci and Stefano Usai,* "Technological activity in European regions"

**05/12** *Rosina Moreno, Raffaele Paci and Stefano Usai,* "Innovation clusters in European regions"

**05/11** *Luca Deidda and Bassam Fattouh,* "Bank, financial markets and growth"

**05/10** *Fabio Cerina and Francesco Pigliaru,* "Agglomeration and growth in NEG: a critical assessment"

**05/09** *Maurice Kugler and Rossella Oppes,* "Collateral and risk-sharing in group lending: evidence from an urban microcredit program"

**05/08** *Adriana Di Liberto,* "Convergence and divergence in neoclassical growth models with human capital"

**05/07** *Maria Gabriela Ladu,* "Growth and employment: a survey on the demand side of the labour market"

**05/06** *Maria Gabriela Ladu,* "Total factor productivity growth and employment: a simultaneous equations model estimate"

**05/05** *Silvia Balia,* "Health and economic behaviour: a critical survey of the literature"

**05/04** *Gianna Boero, Tiziana Laureti and Robin Naylor,* "An econometric analysis of student withdrawal and progression in post-reform Italian universities"

**05/03** *Vittorio Pellagra,* "Banking with sentiments. A model of fiduciari interactions in micro-credit programs"

**05/02** *Margarita Genius and Elisabetta Strazzera,* "Modelinf elicitation effects in contingent valuation studies: a Monte Carlo analysis of the bivariate approach"

**05/01** *Fabio Cerina,* "Marshall's *ceteris paribus* in a dynamic frame"

**04/20** *Rinaldo Brau and Carlo Carraio,* "The economic analysis of voluntary approaches to environmental protection. A survey"

**04/19** *Cristina Murroni,* "ICT and local development: a case study of the metropolitan area of Cagliari"

**04/18** *Adriana Di Liberto,* "Convergence clubs and the role of human capital in Spanish regional growth"

**04/17** *Stefano Usai and Marco Vannini,* "Banking structure and regional economic growth: lessons from Italy"

**04/16** *Silvia Balia and Andrew M. Jones,* "Mortality, lifestyle and socio-economic status"

**04/15** *Rosina Moreno, Raffaele Paci and Stefano Usai,* "Geographical and sectoral clusters of innovation in Europe"

**04/14** *Gianfranco Atzeni and Oliviero Carboni,* "ICT productivity and firm propensity to innovative investment: learining effects evidence from Italian micro data"

**04/13** *Fabrizio Ariani and Luca G. Deidda,* "Few bad apples or plenty of lemons: which makes it harder to market plums?"

**04/12** *Giovanni B. Concu,* "A choice modelling approach to investigate biases in individual and aggregated benefit estimates due to omission of distance"

**04/11** *Giovanni B. Concu,* "Effects of distance aon non-use values"

**04/10** *Antonio Sassu and Sergio Lodde,* "Le tecnologie dell'informazione e della comunicazione nelle piccole imprese dei settori maturi"

**04/09** *Riccardo Marselli e Marco Vannini,* "L'efficienza tecnica dei distretti di Corte d'Appello italiani: aspetti metodologici, benchmarking e arretrato smaltibile"

**04/08** *Rinaldo Brau, Matteo Lippi Bruni and Anna Maria Pinna,* "Public vz private demand for covering long term care expenditures"

**04/07** *Matteo Bellinzas* "Dinamiche demografiche, agglomerazione e determinanti economiche. Il caso italiano"

**04/06** *Giuseppe Medda, Claudio Piga* "R&S e spillover industriali: un'analisi sulle imprese italiane

**04/05** *Adriana Di Liberto, Roberto Mura, Francesco Pigliaru* "How to measure the unobservable: a panel technique for the analysis of TFP convergence"

**04/04** *Vittorio Pelligra* "How to incentive Who? Intr-personal and inter-personal mechanisms"

**04/03** *Domenica Giovanna Dettori, Antonello Paba, Manuela Pulina* "European rural Tourism: agrotouristic firms in Sardina and their life cycle"

**04/02** *Giovanni Sulis* "Wage dispersion and Equilibrium search models: some evidence from Italy"

**04/01** *Silvia Loddo* "Old and new intervention policy: A survey of empirical studies for the Mezzogiorno"

**03/13** *Raffaele Paci, Ernesto Batteta* “Innovation Networks and Knowledge Flows across the European Regions”

**03/12** *Antonio Sassu, Sergio Lodde* “Piccole imprese e tecnologie della comunicazione: un’indagine empirica”

**03/11** *Antonio Sassu, Sergio Lodde* “Tradizione e innovazione nel settore vinicolo in Sardegna”

**03/10** *Rosina Moreno-Serrano, Raffaele Paci, Stefano Usai* “Spatial distribution of innovation activity. The case of European Regions”

**03/09** *Rinaldo Brau, Alessandro Lanza, Francesco Pigliaru* “How fast are the tourism countries growing? The cross-countries evidence.

**03/08** *Margarita Genius, Elisabetta Strazzera* “The copula approach to sample selection modelling: an application to the recreational value of forests”

**03/07** *Guido Ascari, Emanuela Marrocu* “Forecasting inflation: a comparison of linear Phillips curve models and nonlinear time series models”

**03/06** *Alessandro Lanza, Matteo Manera, Massimo Giovannini* “Oil and product price dynamics in international petroleum markets”

**03/05** *Rinaldo Brau, Gianluca Fiorentini, Matteo Lippi Bruni, Anna Maria Pinna* “La disponibilità a pagare per la copertura del rischio di non autosufficienza: analisi econometrica e valutazioni di policy”

**03/04** *Gianfeanco Atzeni, Claudio Piga* “ Credit Rationing in High-Tech Firms and Sample Selection”

**03/03** *Manuela Pulina* “Quantitative forecasting for tourism: OLS and ARIMAX approaches”

**03/02** *Gianna Boero, Riccardo Pinna* “Durata degli studi e voto di laurea: un'indagine econometrica su alcune facoltà dell'Università di Cagliari”

**03/01** *Anna Maria Pinna, Miriam Manchin* **“**Border Effects in the Enlarged EU Area. Evidence from Imports to Applicant Countries”

**02/12** *Gianna Boero, Riccardo Pinna* “La performance esterna dell'Università di Cagliari: il caso delle facoltà del polo giuridico-economico”

**02/11** *Sergio Lodde* “Patti territoriali e specializzazione produttiva nel Mezzogiorno”

**02/10** *Maria Grazia Curreli, Sergio Lodde* ‘Saperi tradizionali e sviluppo locale: il comparto della produzione del miele in Sardegna”

**02/09** *Gianna Boero, Jeremy Smith, Kenneth F. Wallis* “The properties of some goodness-of-fits tests”

**02/08** *Gianna Boero, Emanuela Marrocu* “The performance of SETAR models: a regime conditional evaluation of point, interval and density forecasts”

**02/07** *Anna Maria Pinna, Carla Fancello* "How far do we trade intermediate inputs?"
**02/06** *Sonia Deidda, Raffaele Paci, Stefano Usai* "Spatial Externalities and Local Economic Growth"
**02/05** *Elisabetta Strazzera, Rinaldo Brau, Silvia Balia, Simone Atzeni* "La disponibilità a pagare e le preferenze dei turisti per i siti del Parco Geominerario della Sardegna: il caso di Porto Flavia"
**02/04** *Rinaldo Brau, Massimo Florio* "Privatisations as price reforms: Evaluating consumers' welfare changes in the UK"
**02/03** *Luca Deidda, Fabio Cerina* "Do we need more time for leisure?"
**02/02** *Luca Deidda, Bassam Fattouh* "Concentration in the Banking Industry and Economic Growth"
**02/01** *Raffaele Paci, Silvia Saddi* "Capitale pubblico e produttività nelle regioni italiane"
**01/14** *Gianfranco E. Atzeni, Oliviero A. Carboni* "The Economic Effects of Information Technology: Firm Level Evidence from the Italian case"
**01/13** *Stefano Usai, Raffaele Paci* "Externalities and Local Economic Growth in Manifacturing Industries"
**01/12** Ma*rzio Galeotti, Alessandro Lanza, Matteo Manera* "Rockets and Feathers Revisited: An International Comparison on European Gasoline Markets"
**01/11** *Gianna Boero, Abigail McKnight, Robin Naylor, Jeremy Smith* "Graduates and Graduate Labour Markets in the UK and Italy"
**01/10** *Gianna Boero, Emanuela Marrocu* "Evaluating Non-Linear Models on *Point* and *Interval* Forecasts:. An Application with Exchange rate return."
**01/09** *Robin A. Naylor*, "Firm Profits and the Number of Firms under Unionised Oligopoly"
**01/08** *Robin A. Naylor*, "Industry Profits and Market Size under Bilateral Oligopoly"
**01/07** *Francesco Pigliaru*, "Analisi della convergenza regionale, troppa o troppo poca?"
**01/06** *Elisabetta Strazzera*, "Stima della domanda turistica nel Parco Marino della Maddalena: un'applicazione del metodo della valutazione contingente"
**01/05** *Luca Deidda*, "Financial Institutions' Expertise and Growth Effects of Financial Liberalisation"
**01/04** *Luca Deidda*, "Non Linearity between Finance and Growth"

Finito di stampare nel mese di Maggio 2006
Presso Editoria&Stampa
Zona Industriale Predda Niedda str. n. 10
07100 Sassari