MICROECONOMETRIC ANALYSIS OF RESIDENTIAL WATER DEMAND

Cristina López-Mayán

CEMFI Working Paper No. 0803

April 2008

CEMFI
Casado del Alisal 5; 28014 Madrid
Tel. (34) 914 290 551 Fax (34) 914 291 056
Internet: www.cemfi.es

I wish to thank Manuel Arellano for his great supervision. I also benefited from comments of Guillermo Caruana, Stéphane Bonhomme and Enrique Sentana. Thanks are also due to Graciela Sanromán. I am also very grateful to the company that provided the data. Financial support form Fundación Ramón Areces is gratefully acknowledged. All errors are mine.

CEMFI Working Paper 0803
April 2008

MICROECONOMETRIC ANALYSIS OF RESIDENTIAL WATER DEMAND

Abstract

I present a microeconometric model to analyse residential water demand using panel data. Pricing has an increasing-block structure. Database contains individual consumptions from water meters and lacks further information. Permanent income is treated as an unobservable individual effect determining optimal consumption. I also consider a time-varying shock to connect latent and observed demands. The economic setup gives rise to a random effects model with a nonlinear individual effect. I use likelihood-based indirect inference for estimation. I compute price-elasticities and predict the effects of a tariff change. The methodology can be applied to analyse demands of other goods with increasing tariffs.

JEL Codes: D12, L95, Q25, Q28.
Keywords: Household demand, panel data, kinked budget constraint, indirect inference, Random effects, price-elasticity, utilities, increasing-block pricing.

Cristina López-Mayán
CEMFI
clopezmayan@cemfi.es

1 Introduction

In this paper, I develop a methodology to analyse residential water demand using panel data of individual consumptions. The price schedule of water is characterized by an increasing-two-block tariff. However, the methodology developed in the paper is not water specific. It can be applied to any good with an increasing pricing schedule. Utilities (electricity, gas, telephone,...) are an example of such kind of goods. They usually have blocking tariffs with several fixed costs associated with connection and maintenance. The main advantage of this methodology is it only needs individual consumptions of the good. And those data do not involve surveys. They come from the records that the firm that manages the utility collects from meters.

The kind of demand model I use in the paper is a standard model with a kinked budget constraint. The analysis of demand models with piecewise linear constraints started with Burtless and Hausman (1978) in the context of labour supply. Later, Hewitt and Hanemann (1995) applied it to water context.

In the case of labour supply, information on individual income is usually present in databases. This makes possible to estimate labour supply functions. However, the situation is different for utilities. In general, it is not easy to obtain individual consumptions of water, gas, electricity,... They come from the records collected from individual meters. As consumptions are private information, firms are reluctant to give them. In any case, even if firm agrees to provide the data, they are anonymous, so it is not possible to match them with other data sources to know income. Authors have to use proxies.This is the case in Hewitt and Hanemann (1995). They use the value of the home for tax assesment purposes as a proxy for the household income.

Difficulties to obtain good data are the most important problem to analyse demand of utilities.

The methodology this paper presents avoids those limitations because it allows analysing individual demand by only using the observed consumptions.

The information I have available is a panel of individual water records. Households face an increasing-two-block tariff with a fixed cost. This cost is associated with the consumption of thirty cubic meters and it is charged to all people although they consume below that amount. The economic rationality of an individual implies he will consume thirty or more units but never less because he has to pay for thirty anyway. However, data do not show this: more than fifty percent of observations are characterized by consumptions strictly less than thirty cubic meters. To take into account this fact, I include an opportunity cost in the model that accounts for factors not present in the tariff but that influence the individual water demand such as, for example, the ecological awareness.

The estimated value for the opportunity cost is 0.437 euros per cubic meter, a non negligible amount taking into account that the higher price in the tariff is 0.33 euros per cubic meter.

The empirical strategy I adopt to estimate the model is to treat income as an unobservable individual effect determining optimal demand. It drives both the discrete choice (the optimal block) and the continuous choice (the units of water). This individual effect represents the permanent income and it is unobservable to the econometrician but not to the household.

I also consider a time-varying measurement error that connects optimal and observed demands. It is unobservable to both the econometrician and the household. It is an unanticipated shock that affects the continuous choice (even, if the shock is big enough, it may move the demand to a different block). In general, it implies that the observed demands are not the utility-maximizing locations.

This economic setup gives rise to a random effects model with a nonlinear individual effect because the kinked budget constraint generates a nonlinear optimal demand. This nonlinearity, jointly with the fact that the individual effect enters exponentially in demands, makes difficult to estimate the model. The method I use for estimation is likelihood-based indirect inference using an approximate auxiliary model instead of a reduced form model. So, the auxiliary model will be a simpler version of the structural model.

The validity analysis shows the model fit is good. I use the estimations to compute the cross-sectional distribution of the price-elasticity. Its mean is -0.12, quite inelastic, but households consuming above the kink have a higher mean elasticity (-0.45). Thus, the effect of the fixed cost is to reduce the response to price changes of households consuming in the first block. Finally, I show how the model can be used to predict the effects of introducing a new tariff.

Therefore, the methodology I develop in the paper is useful to understand the individual demand of any good with an increasing-block tariff but it is also useful to make policy analysis.

The rest of the paper is organized as follows. In section 2, I comment briefly the literature. In section 3, I describe the data and comment some preliminary analysis. In section 4, I develop the microeconomic model and section 5 presents the empirical strategy. The results are in section 6 and the policy analysis in section 7. Section 8 presents the conclusions.

2 Overview of the literature

One of the most important characteristics of goods with nonlinear prices is the simultaneous determination of the block (the discrete choice) and the quantity consumed (the continuous choice). Burtless and Hausman (1978) was the first study to deal with this question. It analyses the labour supply in presence of progressive income taxation. In that case, the after-tax wage depends on the number of hours of work supplied and the hours worked is a determinant of the after-tax wage. The budget constraint is not linear but can be convex or concave due to transfer payments.

Since then, several authors have analysed the econometrics of nonlinear budget sets. Some examples are Hausman (1985), Moffitt (1990) and Pudney (1990). In addition, Hanemann (1984) develops a framework to analyse demand models in which the consumer face discrete-continuous choices. The discrete decision depends on the continuous choice and vice versa.

The water demand is a natural context to apply the discrete-continuous approach because the tariffs are usually nonlinear. The consumer has to take two decisions: the block and how much water he will consume within that block.

The objective of the literature of water demand is to estimate the price-elasticity. First studies use regressions to analyse the influence of price in water demand. The question was about whether using as regressor the marginal or the average price. Taylor (1975) and Nordin (1976) were the first who argument that both the marginal price and the difference variable must be included in the regression. The difference variable accounts for the fact that the marginal price may not be the price of every consumed unit.

Since those two works, a lot of studies of water demand include the marginal price and the difference variable as explanatory variables. They estimate the regressions using IV, 2SLS or 3SLS to avoid the endogeneity derived from the simultaneous decision of marginal price and consumption. An example of this studies is Nieswiadomy and Molina (1989).

However, these works left unmodeled the choice of the block in which the individual consumes. Hewitt and Hanemann (1995) apply the discrete-continuous choice approach to water demand. With this framework, the discussion of whether to use marginal or average price disappears because the marginal price enters in the budget constraint. The difference variable is also implicit in the constraint so its effects in the optimal demand are taken into account.

With respect to price-elasticity estimations, in general, papers conclude that water demand is quite inelastic although there are differences depending on the kind of tariff. Nieswiadomy and Molina (1989) use monthly temporal data of a customer sample of Denton (Texas) for the summer months. They obtain elasticities of -0.36 and -0.55 under decreasing and increasing block rates respectively. Hewitt and Hanemann (1995) apply the discrete-continuous approach to the data of Nieswiadomy and Molina (1989) corresponding to the period with increasing block tariffs and obtain elasticities between -1.57 and -1.63.

Dalhuisen, Florax, de Groot and Nijkamp (2001) use meta-analysis techniques to synthesize research results on price and income elasticities of residential water demand. They conclude that the mean elasticity is -0.43 (with a standard deviation of 0.92). Increasing block rate pricing rises price-elasticity but decreasing block rates does not affect its magnitude. In addition, the use of aggregate data instead of household data reduces price-elasticity. They also obtain some evidence that elasticities are higher at long run than at short run.

A recent paper (Olmstead, Hanemann and Stavins 2007) applies the discrete-continuous approach to estimate price-elasticity using daily household consumptions. It uses a very rich dataset corresponding to eleven urban areas in United States and Canada. Households were randomly obtained from customer databases of residential single-family households. Then, selected households were interviewed to obtain sociodemographic variables (annual income, number of residents in household, number of bathrooms,...). Authors also have information on weather conditions. Daily consumptions correspond to four weeks of a year (two weeks for the arid season and two weeks for the wet season). The pricing scheme includes incresing-block tariffs and uniform prices. The price-elasticity for households under increasing-block rates is -0.59 and for households under uniform prices is -0.33. They explore the possibility of endogenous price structures as an explanation for those different price-elasticities.

When household consumption data are used, the availabilty of rich datasets is not usual. The methodology I develop in this paper allows to analyse individual demands when the only available information is household consumptions.

3 Data: descriptive analysis

For this work, I could obtain microdata from a town in the Galician Community, in Spain. Given confidentiality restrictions, from here on, I will call it simply Town.

The Galician Community is in the Northwest of Spain and it is usually a rainy region. Town is in the Northwest of the Community, in an industrialized area. But it is not among the most populated cities in the Galician Community.

The database is a panel of quarterly water consumptions. The data are obtained directly from the household water meters. So, data correspond to the consumption made by all people living in a house[1]. However, the database does not have information about number of members, so my analysis unit will be the household. Variables with information about electrical appliances in households, bathrooms or income are neither available.

In general, water bill includes two issues: one part is associated to the consumed water and another part is associated to the sewer system. However, in the Galician Community, it is very common that bill does not have some of those concepts. There are two main reasons for that fact. First, as the Galician Community is a rainy region, people can obtain water from private wells. So, they do not consume water from the public system and bill does not include the consumption tariff. Second, in the Galician Community houses are disperse over all the

[1] Water meters can be shared by more than one household. The database does not inform about how many households have common meters. However, according to the company, their use is very residual in Town.

territory. In consequence, a lot of them do not have sewer system; thus, they do not pay for it. With the exception of the main towns, these situations are common. Additionally, in many cases, the private company manages only one concept (consumption or sewer system) and the council the other one.

In consequence, in the database I observe households paying for the sewer system and for the consumption, and households paying for only one of those two concepts. On the other hand, the criterion to classify a consumer as domestic varies between the consumption and the sewer tariffs (a household could be domestic for the first but not for the second, although these cases are rare). For all these reasons, I decided to study the demands of households with the two following characteristics: they are considered domestic for the consumption tariff and they pay both for the consumption and for the sewer system.

The database is a quarterly unbalanced panel because households receive water bills quarterly. Observations start in 2001_Q3 and finish in 2005_Q1. Each year has two tariffs, one for the consumption and one for the sewer system. Households that pay both for the consumption and for the sewer face marginal prices equal to the sum of the two tariffs. The tariff is characterized by an increasing-block structure with a fixed cost. The number of blocks is two, with the jump in thirty cubic meters, until year 2004 (see table 1). The fixed cost implies that all people consuming thirty or less cubic meters have to pay the price corresponding to thirty although they consume less than that quantity. The tariff since 2005 has four blocks but the fixed cost disappears and people only have to pay for what they consume.

Table 2 describes the characteristics of the panel until 2004. Town has 52432 observations and 4394 households. 74.56% of the households are a balanced subpanel.

The database has 5786 observations with consumption equal to zero. They represent 11% of total observations and are distributed in a similar way among quarters (first and second quarters have 22% of the zeros, the third quarter 26% and the fourth one 30%). This indicates they are not the result of an important seasonal behaviour. So, I can drop them from the database and work with log-data. After eliminating those consumptions, the base remains with 46646 observations and 4266 households, 54.34% are the balanced subpanel (table 3).

Table 4 shows the percentage of observations at each block and at the kink. More than fifty percent of observations are below thirty cubic meters and only 2.4% have a consumption equal to thirty. However, close to the kink there are more observations.

Table 5 reflects some descriptive statistics of the distribution shown in figure 1. We can see that the median is below the logarithm of thirty what implies that more than fifty percent of households are at the first block. The skewness coefficient goes in the same direction: it is negative which indicates there is more probability on the left hand side of the distribution. One

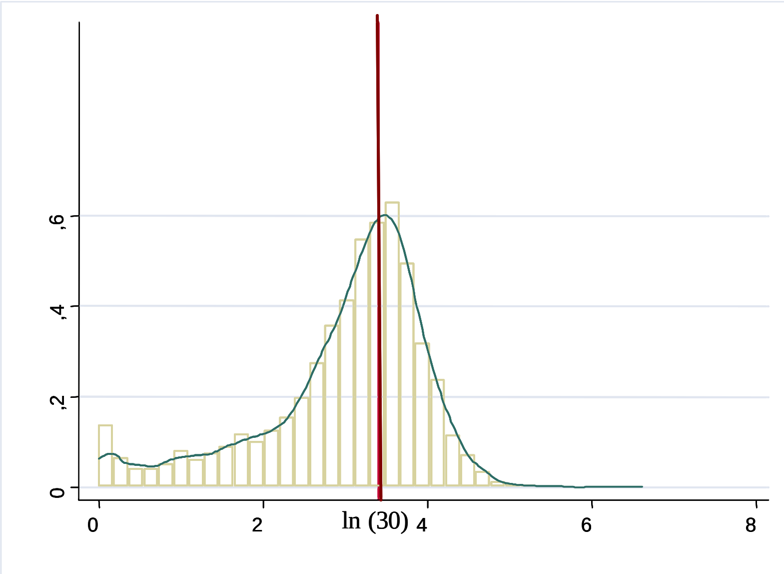

Figure 1: Histogram of cross-sectional distribution of observed log-demands. The distribution is obtained by averaging quarterly log-consumptions for each household. The database size is 46426 observations and it excludes individuals with one observation.

of the most important things in figure 1 is that there are households consuming strictly less than thirty cubic meters.

Another interesting aspect to analyse is the level of block mobility, that is, to which extent households consume always in the same block. According to table 6, the level of mobility is high. Around 52% of the households are not at the same block all quarters whereas 28% of them consume below the kink all quarters and about 19% consume above it.

In the panel, the household average consumption is 33.22 m^3 by quarter and his standard deviation is 46.94 m^3. It is important to analyse what part of the data variability is permanent and what part is transitory because if the latter was predominant, the predictions of the model would not be useful.

The decomposition of the data into their variation sources is obtained by estimating a random effects model with seasonal dummies. This regression allows calculating the variability sources of log-consumption when the influence of seasonality is eliminated. Equation (1) shows the regression: D is the vector of seasonal dummies, η_i is the individual effect and ε_{it} is an *iid* transitory error with normal distribution $N(0, \sigma_\varepsilon^2)$.

$$\log m_{it}^3 = D'\gamma + \eta_i + \varepsilon_{it} \tag{1}$$

Table 7 presents the estimations. The coefficients of the dummies are expressed in terms of the first quarter and are always significant. The presence of seasonality is not surprising in the case of water.

Table 8 contains the variances from the random effects model (1). Data vary mainly for permanent facts: the variance of the individual effect is around 79% of the total data variation.

To measure more in detail the importance of the temporal component in data, I calculate the variance of the individual log-consumptions in deviation to the temporal average. For each year and quarter, I average individual consumptions and, then, I calculate the difference between individual data and the corresponding temporal average. This transformation removes all the temporal component of the water consumed by the households. The variance of this series is 1.07 and it explains 76.76% of the variance of the log-consumptions.

Finally, I analyse the persistence of the shock to know if it has effects more than one period. So, I regress the log-consumption in deviation to temporal average, $(\log m^3_{it})_D$, on lags of this variable. I include five lags to avoid autocorrelation in the error[2]:

$$\begin{aligned}(\log m^3_{it})_D = {} & \alpha_1(\log m^3_{it-1})_D + \alpha_2(\log m^3_{it-2})_D + \alpha_3(\log m^3_{it-3})_D \\ & + \alpha_4(\log m^3_{it-4})_D + \alpha_5(\log m^3_{it-5})_D + \eta_i + u_{it}\end{aligned} \quad (2)$$

where η_i is the individual effect and u_{it} is $iid \sim N(0, \sigma^2_u)$. Estimations are in table 9.

The effect of the shock is persistent. With the exception of lag two, the other lags are significant with the greater value corresponding to lag one. The shock has persistent impact on water consumption specially at short term.

4 The microeconomic model

The microeconomic water demand model presents a kinked budget constraint as a consequence of the increasing-block prices.

The tariff of Town until 2004 has two blocks with a kink in thirty cubic meters. Marginal prices for each block are 0.40 and 0.48, respectively. Therefore, the budget constraint will have two segments.

The tariff also includes an implicit fixed cost because all people have to pay the price of consuming thirty cubic meters although they consume less than thirty. I denote by FC this fixed cost:

$$FC = 30 \ \times \ 0.40 \ eur/m^3 = 12.13 \ euros/quarter \quad (3)$$

The first section of the budget constraint is flat since the relevant marginal price for a consumer that optimizes there is zero, instead of 0.40, due to FC. In the second section, the relevant marginal price (and the slope) is 0.48.

[2] In a quarterly seasonal ARIMA, $(1-\theta_1 L)(1-\theta_4 L^4)w_t = u_t$, it is usual to include lags 1, 4 and 5. In the final regression, I also include lags 2 and 3 because they were significant in previous regressions.

In a constraint with those characteristics, there are two possible optima: the kink, for all agents that choose to consume in the flat section, and the tangency value between the indifference curve and the second branch of the constraint for those that optimize there (see figure 2). This implies there should not be consumptions below thirty cubic meters and many observations equal to thirty.

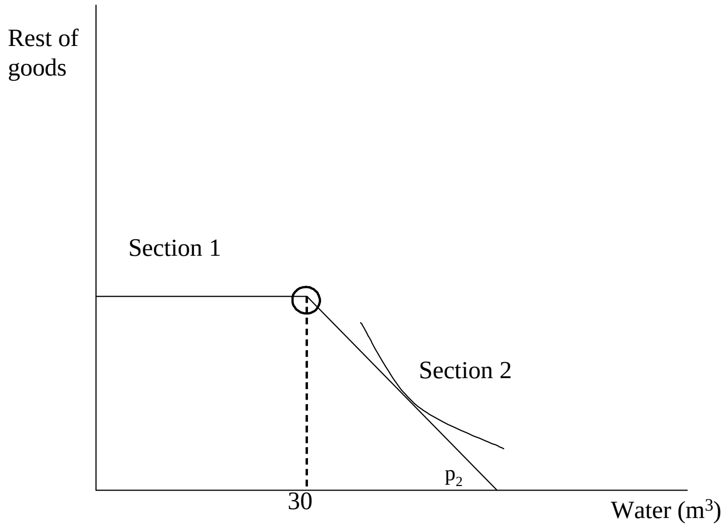

Figure 2: Budget constraint and possible optima. p_2 is the marginal price in block 2. For people consuming at section 1 the optimum is thirty (the kink) and for people consuming at section 2 the optimum is the tangency point.

The data analysis of the previous section does not confirm those predictions. According to table 4, there are 54% of observations below thirty and only 2.4% exactly equal to thirty. So, a budget constraint like the one shown in figure 2 does not explain the data.

Until this point, I have considered that the relevant price for a consumer in the first block was zero because I have only taken into account the tariff. But it is reasonable thinking about there are other factors, different from prices, that can influence individual consumption decisions. These factors could explain why there are so many observations below the kink.

One way of considering those factors is by mean of a *water opportunity cost* (c). This cost will include several aspects such as the ecological awareness of the agent or the cost of controlling if he has already reached the kink. The first one implies an agent will not use thirty cubic meters if he does not need them in spite of having to pay the fixed cost. The second one is explained by the fact that the consumption above thirty will be charged to a higher marginal price; so, household has to control when he reaches the kink in order "to close the tap" because, otherwise, he would pay a higher price. The opportunity cost can include whatever reason different from the official tariff that helps to understand why the agent decides to consume below the kink.

This opportunity cost implies a slope equal to c for the first part of the constraint because now c is the relevant marginal price. Thus, it is possible to optimize in any of its points (see figure 3). I assume that c is present in the two blocks of the tariff (so, the relevant marginal price in the second branch is $c + p_2$) and that it is constant over time and among individuals.

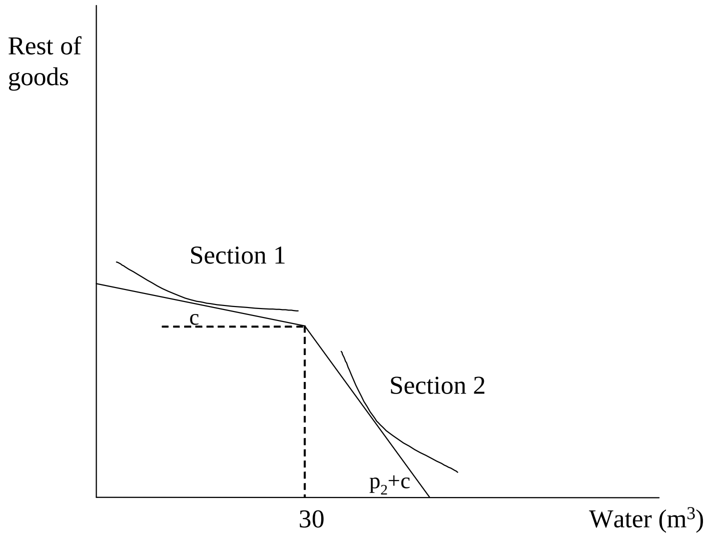

Figure 3: Budget constraint with the opportunity cost. c is present in the two blocks so, slopes in section 1 and 2 are c and $c + p_2$ respectively. Possible optima are tangency points in both sections of the budget constraint.

I assume a representative agent optimizes a Cobb-Douglas utility function subject to the following budget constraint:

$$\begin{aligned} \underset{\{rg,\, a\}}{Max} \quad u(rg, a) &= (rg)^{1-\nu} a^{\nu} \\ st. \qquad m &= rg + FC + ca \qquad\qquad\qquad if \quad a \leq 30 \\ m &= rg + FC + ca + p_2(a - 30) \quad if \quad a > 30 \end{aligned} \tag{4}$$

where a is water consumption, rg is consumption of rest of goods (it is the numerary), ν is the Cob-Douglas coefficient and represents the weight of water consumption in individual utility, m is income and p_2 is the price of the second block.

The second segment of the budget constraint can be rewritten in the following way:

$$m + d = rg + ca + p_2 a \tag{5}$$

where d is equal to $30(p_2 - p_1)$. It reflects an implicit income for those that optimize in the second block. d appears because individuals that consume more than thirty do not pay p_2 for every unit. That income is called *difference variable* and it accounts for the lump sum tranfers implied by block rates.

Some authors defend that the difference variable benefits only second block consumers, so they consider the increasing tariff is inefficient to promote water saving. However, if the highest price of the tariff is efficient from an economic and social point of view, the group of larger consumption is paying the correct price for, at least, some units. For the rest of the consumption, they benefit from the implicit subsidy of the tariff. But, at the same time, this subsidy can allow charging a smaller price to individuals consuming in the first block who usually will have a lower income. Thus, although they do not pay the efficient price, they can access to the water vital minimum.

Even in the case the higher price is not efficient, an increasing tariff generates income transfers among individuals consuming at different blocks. These transfers benefit people consuming in the first block for the reason explained above and also benefit people at higher blocks because they pay a price smaller than the efficient. The problem is not the increasing tariff itself but the prices. Prices of the last blocks should be near to the efficient price.

The first order conditions of problem (4) are:

(6) $\frac{1-\nu}{\nu}\frac{a}{rg} = \frac{1}{c}$, for the first block

(7) $\frac{1-\nu}{\nu}\frac{a}{rg} = \frac{1}{c+p_2}$, for the second block.

Combining these expressions with the suitable branch of the budget constraint, I obtain the optimal water demands:

Demand at section 1: (8) $a_1(m) = \frac{\nu}{c}(m - FC)$

Demand at section 2: (9) $a_2(m) = \frac{\nu}{c+p_2}(m - FC + 30p_2)$

Finally, the optimal water demand function has three branches: first and third are (8) and (9). The second one (demand at the kink) appears when optimization in the other branches is at unfeasible points. This is showed in figure 4. Let us assume that if an individual optimizes in the first blcok, he chooses a point like E and if he optimizes in the second block, he decides a point like D. As the budget constraint is ABC line, neither E nor D is feasible. Thus, the kink is the optimum[3].

Expression (10) resumes the water demand function :

$$a = \begin{cases} \frac{\nu}{c}(m - FC) & if\ a_1(m) < 30 \\ 30 & if\ a_1(m) \geq 30\ and\ a_2(m) \leq 30 \\ \frac{\nu}{c+p_2}(m - FC + 30p_2) & if\ a_2(m) > 30 \end{cases} \quad (10)$$

[3] Hausman (1979) shows that when the budget set is convex and an interior solution is obtained either at a kink or at a tangency with a budget segment, then that optimum is unique.

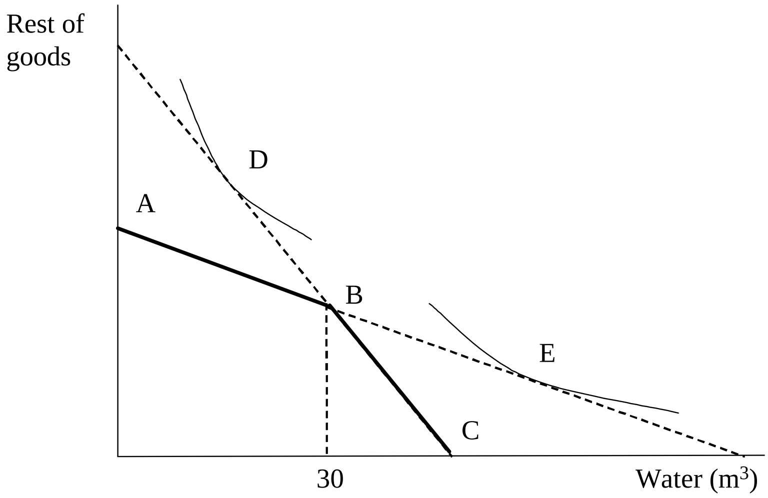

Figure 4: Demand at the kink. The optimum is the kink if a household chooses points like E or D when he optimizes in the first or second block respectively.

5 Empirical strategy

In this section, I explain the methodology to estimate the microeconomic model.

The first step is to linearize expressions (8) and (9):

(11) $\log a = \log \nu - \log c + \log (m - FC)$

(12) $\log a = \log \nu - \log(c + p_2) + \log (m - FC + 30p_2)$

m is interpreted like the part of the individual permanent income corresponding to the sample period. It presents cross-sectional but not temporal variability. This is not a restriction because the panel is short by the temporal side and, so, it is feasible to think there are not important changes in individual permanent income over the sample period. However, the cross-sectional variability is important because the panel has many individuals.

Consequently, this permanent income represents an individual effect whose variability explains the most of the water consumption variation. An important part of this consumption comes from electrical appliances, number of bathrooms,..., goods that depend on individual income positively.

I assume m has an exponential form, $m_i = e^{\beta_i}$, where $i = 1, ..., N$ indicates an agent and β_i is the individual effect on water demand. It is unobservable to the econometrician but not to the household.

I rewrite expressions (11) and (12) to incorporate the previous assumption:

(13) $\log a_i = \log \nu - \log c + \log \left(e^{\beta_i} - FC\right)$

(14) $\log a_i = \log \nu - \log(c + p_2) + \log \left(e^{\beta_i} - FC + 30p_2\right)$

Finally, the linearized individual water demand is characterized by the following expression:

$$\log a_i = \begin{cases} \log\nu - \log c + \log\left(e^{\beta_i} - FC\right) \\ \quad if \quad \beta_i < \log\left(\frac{30c}{\nu} + FC\right) \\ \\ \log 30 \\ \quad if \;\; \log\left(\frac{30c}{\nu} + FC\right) < \beta_i < \log\left(\frac{30(c+p_2)}{\nu} + FC - 30p_2\right) \\ \\ \log\nu - \log(c+p_2) + \log\left(e^{\beta_i} - FC + 30p_2\right) \\ \quad if \quad \beta_i > \log\left(\frac{30(c+p_2)}{\nu} + FC - 30p_2\right) \end{cases} \tag{15}$$

In (15), the only variable that generates individual variation is β_i. Two people with the same permanent income will have the same optimal water demand. On the other hand, the individual optimal decision will be the same over time because individual permanent income is constant.

The individual effect drives both the discrete choice (the optimal block) and the continuous choice (the units of water).

Under the assumption $\beta_i \sim N\left(\mu_\beta, \sigma_\beta^2\right)$, the probability of consuming at each segment is:

1. $\log a_i = \log\nu - \log c + \log\left(e^{\beta_i} - FC\right)$ with probability $p_1 = p\left[\beta_i < \log\left(\frac{30c}{\nu} + FC\right)\right] = \Phi\left[\frac{\log\left(\frac{30c}{\nu}+FC\right)-\mu_\beta}{\sigma_\beta}\right]$.

2. $\log a_i = \log\nu - \log(c + p_2) + \log\left(e^{\beta_i} - FC + 30p_2\right)$ with probability $p_3 = p\left[\beta_i > \log\left(\frac{30(c+p_2)}{\nu} + FC - 30p_2\right)\right] = \Phi\left[\frac{-\log\left(\frac{30(c+p_2)}{\nu}+FC-30p_2\right)+\mu_\beta}{\sigma_\beta}\right]$.

3. And $\log a_i = \log 30$ with probability $p_2 = 1 - p_1 - p_3 = \Phi\left[\frac{\log\left(\frac{30(c+p_2)}{\nu}+FC-30p_2\right)-\mu_\beta}{\sigma_\beta}\right] - \Phi\left[\frac{\log\left(\frac{30c}{\nu}+FC\right)-\mu_\beta}{\sigma_\beta}\right]$.

β_i is a continuous variable, so the probability of $\log a_i = \log 30$ is zero. However, the model predicts an individual can choose to consume thirty cubic meters. So, I determine the interval of β_i values for what occurs this and I take the probability of this interval like the probability of optimizing at the kink. Thus, the water demand distribution has a point that accumulates probability.

Finally, I incorporate into the model a time-varying term, $\varepsilon_{it} \sim iid\ N\left(0, \sigma_\varepsilon^2\right)$, that connects optimal and observed demands and that is assumed to be independent of β_i. ε_{it} is unobservable to both the econometrician and the household. It is an unanticipated shock that affects the continuous choice[4]. With this variable, the individual final consumption will be different from one quarter to another. It also implies that an individual observed at any segment of the constraint may in fact have an optimum on a different block. In general, the effect of the shock is that observed demands are not utility-maximizing locations.

Despite the kink is a possible optimum, the introduction of ε_{it} implies that there will not be observations equal to thirty cubic meters. In table 4 we can see that, really, few data are at the kink but closer to it there are quite observations. This can be reflecting the fact that an agent chooses to consume at the kink but, then, the shock realizes and moves his consumption a little below or above thirty.

Therefore, observed data y_{it} are the result of combining optimal demand and error:

$$\log y_{it} = \begin{cases} \log a_i^1 + \varepsilon_{it} \quad if \quad \beta_i < \log\left(\frac{30c}{\nu} + FC\right) \\ \log a_i^k + \varepsilon_{it} \quad if \ \log\left(\frac{30c}{\nu} + FC\right) < \beta_i < \log\left(\frac{30(c+p_2)}{\nu} + FC - 30p_2\right) \\ \log a_i^2 + \varepsilon_{it} \quad if \quad \beta_i > \log\left(\frac{30(c+p_2)}{\nu} + FC - 30p_2\right) \end{cases} \tag{16}$$

where $t = 1, ..., T_i$ indicates quarter, $\log a_i^1 \equiv \log \nu - \log c + \log\left(e^{\beta_i} - FC\right)$, $\log a_i^k \equiv \log 30$ and $\log a_i^2 \equiv \log \nu - \log(c + p_2) + \log\left(e^{\beta_i} - FC + 30p_2\right)$.

Expression (16) corresponds to a random effects model with a nonlinear individual effect.

Let $\log y_i = (\log y_{i1}, \log y_{i2}, ..., \log y_{iT_i})$ be the vector of temporal observations for an individual. Then, his likelihood is equal to:

$$f(\log y_i) = f(\log y_{i1}, \log y_{i2}, ..., \log y_{iT_i}) = \int_{-\infty}^{+\infty} f(\log y_{i1}, \log y_{i2}, ..., \log y_{iT_i}/\beta_i) f(\beta_i) d\beta_i \tag{17}$$

This integral can be expressed as the sum of three terms, one for each branch of the log-water demand:

$$\begin{aligned} f(\log y_{i1}, \log y_{i2}, ..., \log y_{iT_i}) &= \int_{-\infty}^{l_1} f(\log y_{i1}, \log y_{i2}, ..., \log y_{iT_i}/\beta_i) f(\beta_i) d\beta_i \\ &+ \int_{l_1}^{l_2} f(\log y_{i1}, \log y_{i2}, ..., \log y_{iT_i}/\beta_i) f(\beta_i) d\beta_i \\ &+ \int_{l_2}^{+\infty} f(\log y_{i1}, \log y_{i2}, ..., \log y_{iT_i}/\beta_i) f(\beta_i) d\beta_i \end{aligned} \tag{18}$$

where $l_1 = \log\left(\frac{30c}{\nu} + FC\right)$ and $l_2 = \log\left(\frac{30(c+p_2)}{\nu} + FC - 30p_2\right)$.

[4] I do not use expression "measurement error" for ε_{it} because it can be confusing. "Measurment error" seems to indicate the econometrician can not measure correctly the consumption. This is not the case given that data come from household meters.

So, the total log-likelihood is $L(\log y_1, \log y_2, ..., \log y_N) = \sum_{i=1}^{N} \log f(\log y_i)$, where f is the density function of individual log-consumptions.

The parameters of interest are c, ν, σ_β^2 and σ_ε^2. I impose some restrictions on them:

$R1$) $c > 0$ and $\nu > 0$ to ensure that the log-demand is well defined.

$R2$) $\nu \leq 1$ to ensure that $l_1 < l_2$.

μ_β is not a parameter to estimate. $m_i \sim \log N\left(\mu_\beta, \sigma_\beta^2\right)$, so, $\mu_\beta = \log E\left(m_i\right) - \frac{1}{2}\sigma_\beta^2$. And I fix $E\left(m_i\right)$ equal to 6725 euros, which is the quarterly available income by a household in 2001[5].

It is not obvious what kind of distributions follow both $\log a_i$ and $\log y_{it}$ because the individual effect, β_i, does not enter linearly in the demand equation. In consequence, the above integrals do not have an analytical expression. The maximization of the log-likelihood would require the use of numerical integrals.

However, there is an alternative methodology that avoids the computional problems of using numerical integrals. This is the likelihood-based indirect inference method. Loosely speaking, it consists of estimating an auxiliary model, simpler than the true one (called structural model), but that captures the main aspects of the latter. The number of parameters in the auxiliary model must be equal or greater than the number of structural parameters. Then, giving values to the vector of structural parameters, data are simulated from the structural model and are used to estimate the parameters from the auxiliary model by maximizing its likelihood.

After M simulations, auxiliary estimations are averaged. This average is a statistic that contains information about the structural parameters. If the exact relationship between the auxiliary and the structural parameters was known, it would be enough to invert it and to obtain an estimator of the structural parameters. In general, it is not possible to know that relationship. The indirect inference method gives a numerical approximation of it.

Once the auxiliary estimations are averaged, this statistic is used to maximize the likelihood of the auxiliary model but now with the true data. The result of the maximization is an estimation of the structural parameters. So, with this method, it is possible to recover the structural parameters without estimating directly the structural model.

In the indirect inference method, the most important is to choose an adequate auxiliary model. As it was said before, it must capture the main aspects of the model we want to estimate and it must have at least so many parameters as the structural model.

In my case, the auxiliary model is an approximation of the structural model. Changing only one assumption, I obtain a model very similar to the structural one, with the same parameters (c, ν, σ_β^2 and σ_ε^2), but that can be estimated without using numerical integrals.

[5] I calculate this quantity using the available income and the number of households in 2001 in the Galician province in which Town is. These data come from the 2001 Census and from the 2001 Spanish Regional Accounts, both available in the Spanish Statistical Institute web page.

The change in the assumption is the following: in expressions (11) and (12) the individual effect is the net income, mn_i instead of m_i.

(19) $\log a = \log \nu - \log c + \log \underbrace{(m - FC)}_{mn}$

(20) $\log a = \log \nu - \log(c + p_2) + \log \underbrace{(m - FC + 30p_2)}_{mn}$

The net income is also exponential, $mn_i = e^{\beta_i}$, where $\beta_i \sim N\left(\mu_\beta, \sigma_\beta^2\right)$. By substituting in (19) and (20),

(21) $\log a_i = \log \nu - \log c + \beta_i$

(22) $\log a_i = \log \nu - \log(c + p_2) + \beta_i$

And the linearized water demand is:

$$\log a_i = \begin{cases} \log \nu - \log c + \beta_i \\ \quad if \;\; \beta_i < \log 30 + \log c - \log \nu \\ \\ \log 30 \\ \quad if \;\; \log 30 + \log c - \log \nu < \beta_i < \log 30 + \log(c + p_2) - \log \nu \\ \\ \log \nu - \log(c + p_2) + \beta_i \\ \quad if \;\; \beta_i > \log 30 + \log(c + p_2) - \log \nu \end{cases} \tag{23}$$

As before, $\log y_{it} = \log a_i + \varepsilon_{it}$ and $\varepsilon_{it} \sim iid\ N\left(0, \sigma_\varepsilon^2\right)$. Because the individual effect enters linearly in the demand, integrals in (18) have analitical expressions[6].

The next step is to simulate a vector of observations for the N individuals using the structural model. This vector depends on the structural parameters:

$$\log y^m(\alpha) = [\log y_1(\alpha), \log y_2(\alpha), ..., \log y_N(\alpha)]$$

where m indicates simulation and α is the vector of structural parameters.

Next, I obtain

$$\widetilde{\theta}^m(\alpha) = \arg\max_\theta L^{aux}\left[\theta; \log y^m(\alpha)\right]$$

where θ is the vector of auxiliary parameters.

After repeating the process for $m = 1, 2, ...M$, I can calculate $\widetilde{\theta}(\alpha) = \frac{1}{M}\sum_{m=1}^{M} \widetilde{\theta}^m(\alpha)$. This statistic contains information about structural parameters. Finally, the estimation of α is:

$$\widehat{\alpha} = \arg\max_\alpha L^{aux}\left[\widetilde{\theta}(\alpha); \log y\right]$$

where $\log y = (\log y_1, \log y_2, ..., \log y_N)$ is the vector of true data.

[6] See appendix 1 for a complete development of the auxiliary model.

It is necessary to note that $\log y_{it}$ is a discontinuous function of the parameters. So, simulated observations are also a discontinuous function of α. And this discontinuity also affects to $\widetilde{\theta}^{m}(\alpha)$, $\widetilde{\theta}(\alpha)$, and, in consequence, to $\widehat{\alpha}$ making not possible to use gradient-based methods of optimization. To avoid this problem, I smooth simulated data following the method proposed by Keane and Smith (2003). For convenience, I express (16) as the following sum:

$$\log y_{it} = \left(\log a_i^1 + \varepsilon_{it}\right) 1\left(\beta_i < l_1\right) + \left(\log a_i^k + \varepsilon_{it}\right) 1\left(l_1 \leq \beta_i \leq l_2\right) + \left(\log a_i^2 + \varepsilon_{it}\right) 1\left(\beta_i > l_2\right) \quad (24)$$

And smoothed data are:

$$\begin{aligned} \log y_{it}^s = & \left(\log a_i^1 + \varepsilon_{it}\right) \Lambda\left(\frac{l_1 - \beta_i}{\lambda}\right) + \left(\log a_i^k + \varepsilon_{it}\right)\left[\Lambda\left(\frac{l_2 - \beta_i}{\lambda}\right) - \Lambda\left(\frac{l_1 - \beta_i}{\lambda}\right)\right] \\ & + \left(\log a_i^2 + \varepsilon_{it}\right)\left[1 - \Lambda\left(\frac{l_2 - \beta_i}{\lambda}\right)\right] \end{aligned} \quad (25)$$

where $\Lambda(r) = \frac{e^r}{1+e^r}$.

The vector of simulated observations, $\log y^m(\alpha)$, is computed using (25). As we can see, when $\lambda \to 0$, $\log y_{it}^s \to \log y_{it}$.

6 Results

The final size of the database I use to estimate the model is 46426 observations (4046 households) because I do not consider 220 individuals that appear only one quarter (they do not contribute to the likelihood).

Table 10 presents the estimations. They were obtained using the indirect inference method with $M = 4$ and $\lambda = 0.05$. To impose restrictions $R1$ and $R2$, I reparametrize $c = e^{c'}$ and $\nu = \frac{e^{\nu'}}{1+e^{\nu'}}$. All the parameters are significant at the 1% level.

The estimated values for σ_β^2 and σ_ε^2 are very similar to those from table 8. In the case of $\widehat{\sigma}_\beta^2$, it is because the individual effect of (1) is a realization of the $\log a_i$ distribution. And the variance of this distribution determines, jointly with the rest of the parameters, the individual effect variance from equation (1). Thus, the value of σ_η^2 depends on σ_β^2, c and ν.

On the other hand, $\widehat{\sigma}_\varepsilon^2$, is equal to the variance of the transitory variable in the error components model of equation (1). ε_{it} is the same error in both cases because I incorporate it into the model after a realization of $\log a_i$. Therefore, its estimated variance must be also the same.

In addition, $\widehat{\sigma}_\beta^2$ is greater than $\widehat{\sigma}_\varepsilon^2$, so the most of the variation in observed consumptions comes from the individual effect not from the unanticipated shock.

The value for ν is very small (0.004). Its interpretation is the individual obtains little utility from the water consumption. The most of utility comes from the consumption of all other goods.

The most important result from table 10 is about the water opportunity cost. c is a structural parameter incorporated into the model due to the necessity of explaining the existence of consumptions below the kink in spite of the presence of a fixed cost. This parameter represents an implicit cost in the tariff that the agent considers when he decides his optimal consumption. Its value is 0.583. It is important to note that it is greater than p_2. Therefore, when the agent maximizes his utility takes into account the fixed cost, the price in the second block and also an additional price, c, not in the tariff.

6.1 Price-elasticity

When price is not linear, there can be different definitions of price-elasticity. One posibility is to define elasticity as a measure of the change in the demand after a one percent change in all marginal prices. Another definition of price-elasticity considers changes in only one price.

The definition I use now is the first one. So, this elasticity measures the response of the individual water demand when p_1 and p_2 increase a one percent. To compute it, I simulate 464260 demands under initial and final prices to calculate the difference between them weighted by the inverse of 0.01. Finally, I average to obtain the distribution of the elasticities across individuals.

The mean of this distribution is -0.12 with a standard deviation of 0.19. It is a very bi-polarized distribution and the reason is the increasing-block tariff. Depending on the block people are, they respond differently to a one percent change in prices.

In consequence, it is more interesting to show separately the elasticities of the agents consuming at block one from the elasticities of those consuming at block two. This is what figure 5 shows.

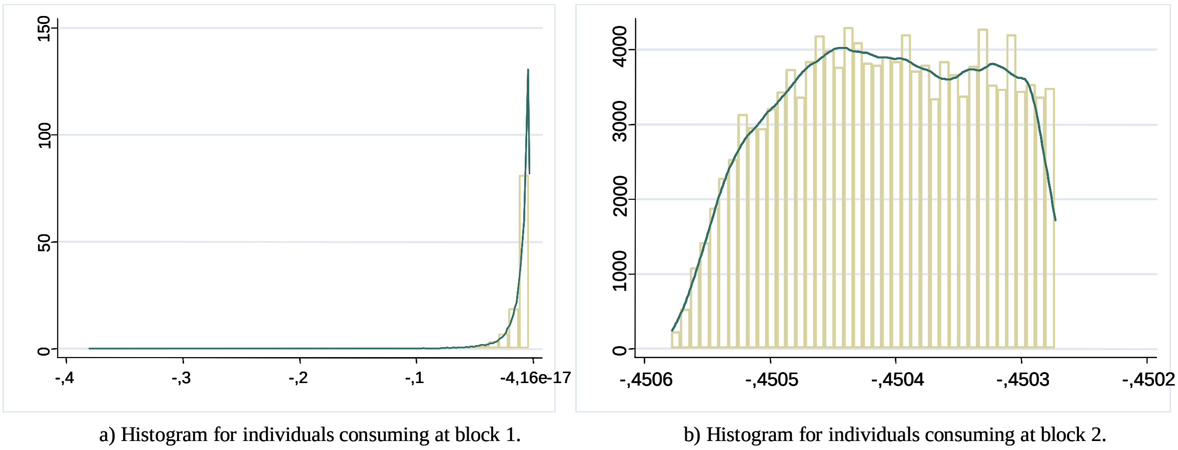

Figure 5: Cross-sectional price-elasticity distributions. These distributions are obtained by simulating 464260 demands under initial prices and under a simultaneous 1% increase in prices. Individual price-elasticity is the temporal average of the difference between both demands weighted by the inverse of 0.01.

Table 11 reports some statistics of the two distributions. The mean elasticity is very different depending on the block at which the household is. Households consuming at block two have a higher mean price-elasticity (-0.45) than households consuming at block one (-0.01). And the distribution of the former households has much less dispersion.

The higher elasticity of households consuming at block two is not surprising. These individuals are more able to adjust their consumptions when prices change because, in general, they are consuming water above the vital minimum.

On the other hand, the lower elasticity of people consuming at block one is explained by several factors. One is they are consuming near to the vital minimum, so their ability to adjust the demand is lower. Another factor is the increase in p_1 affects them independently of how much water they are consuming (their relevant price is c which does not change). For these reasons the change in the demand of those consuming at block one is so small.

As I remarked previously, when the price is not linear, there is an alternative definition of the elasticity based on the change of only one price. I also compute price-elasticities when only p_1 or p_2 increases one percent.

If p_1 increases, both people consuming at block one and at block two respond. The mean elasticity of those at block one is -0.01, like in the case in which both prices change. And the mean elasticity of those at block two is -0.001 (very small due to p_1 affects demand by mean of a fixed cost).

If p_2 increases, only households at block two change their demands. Their mean elasticity is -0.45, like when p_1 and p_2 change simultaneously.

Therefore, I reach the same results than with the first definition of elasticity. The reason is that a one percent increase in the two prices is so small that almost nobody jumps to another block. So, really, everybody responds as if only the price of his block changes.

The price-elasticities obtained in this paper are very similar to those estimated by Martínez-Espiñeira (2002). He calculates price-elasticity using data of several cities of the Northwest of Spain. So, elasticities are estimated using data from the same Spanish region as the data I use in this paper. In addition, the tariffs of these cities have the same characteristics than the tariff of Town. They are increasing-block tariffs with the relevant marginal price in the first block equal to zero (that is, with an implicit fixed cost). The range of price-elasticities Martínez-Espiñeira (2002) obtains is between -0.12 and -0.16, and the overall price-elasticity I compute (-0.12) lies inside. Although his data are at the municipal level, his estimations are a good reference framework for my results.

Additionally, I observe a higher elasticity (-0.45) for those households consuming at block two. Martínez-Espiñeira (2002) also concludes that people consuming above the first block are more responsivness to a change in the marginal price, mainly due to the presence of the fixed cost. The elasticity he estimates for those people is around -0.3, some lower than the elasticity I obtain.

The magnitude of the overall mean elasticity confirms the consensus that water demand is quite inelastic. However, this mean hides important differences that appear when people is looked by consumption blocks.

6.2 Model validity

Here I present some results to evaluate the validity of the microeconometric model.

The first column from table 12 indicates the percentage of observations with consumption lower, equal or higher than thirty cubic meters. Second and third columns present the predicted probabilities. The second column contains the probabilities before the shock ε_{it} realizes and column three has the probabilities after happening ε_{it}[7]. The more highlighting difference between these columns is in the kink probability. When an agent solves his utility maximization problem, it is perfectly possible that the optimum was thirty cubic meters. In the column two, the probability of being at the kink is 0.18, clearly different from zero. However, in the column three, it is zero. And this is so because, after the shock realizes, the probability of consuming thirty is always zero because ε_{it} is a continuous variable. This allows reproducing the observed fact that there are few observations at the kink.

The percentages from data have to be compared with those from the last column (once the shock happens). From this comparison, it can be concluded that model is able to reproduce data behaviour.

Figure 6 shows the predicted cross-sectional distribution of log-consumptions. I obtain it by simulating 464260 individual demands (40460 households) from the model replacing the parameters by their estimations. Then, I compute the average of the observations for each individual.

Additionally, table 13 compares some statistics of that distribution with those from the distribution in figure 1. The mean is almost equal in both distributions and the median is some smaller than the median of the data distribution. The percentiles near to the median are well predicted and those in the tails of the distribution are predicted some worse. But, in general, the simulated distribution is similar to the true one.

[7] The probabilities before the shock are calculated using expressions developed in section 5. Appendix 3 contains expressions for probabilities after realizing ε_{it}.

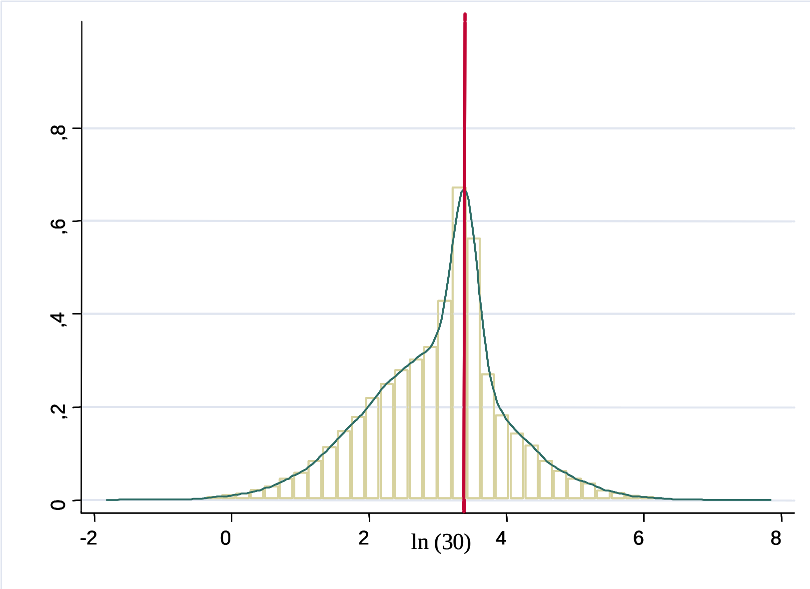

Figure 6: Histogram of cross-sectional distribution of simulated log-demands. It is obtained by simulating 464260 observations (40460 households) and then averaging them across individuals.

7 Policy analysis

In this section, I use the model to obtain policy implications. For example, the model can be useful to know how the household water demand evolves when prices change. To show that, I simulate 464260 demands under different increases in prices and, by averaging them across individuals, I obtain a collection of cross-sectional distributions that show the evolution of the individual demands.

I consider both the effects a rise in the two prices and an increase in only one price. Table 14 shows what changes I consider.

Figure 7 contains the histograms corresponding to different simultaneous increases in p_1 and p_2. Table 15 presents some statistics from those distributions.

As the prices increase, there is more density on the left of the distribution. The median and the rest of percentiles have lower values, the skewness is more negative and also the dispersion reduces. That is, as prices increase, all people reduce their demands, especially those who are in the second block. They accumulate around the kink. However, the part of the distribution on the right of the kink does not disappear even when prices increase a lot (200%). This indicates some people have water necessities that force them to demand above thirty cubic meters independently of how much prices increase.

I also analyse how the cross-sectional distributions behave when only one price varies. The

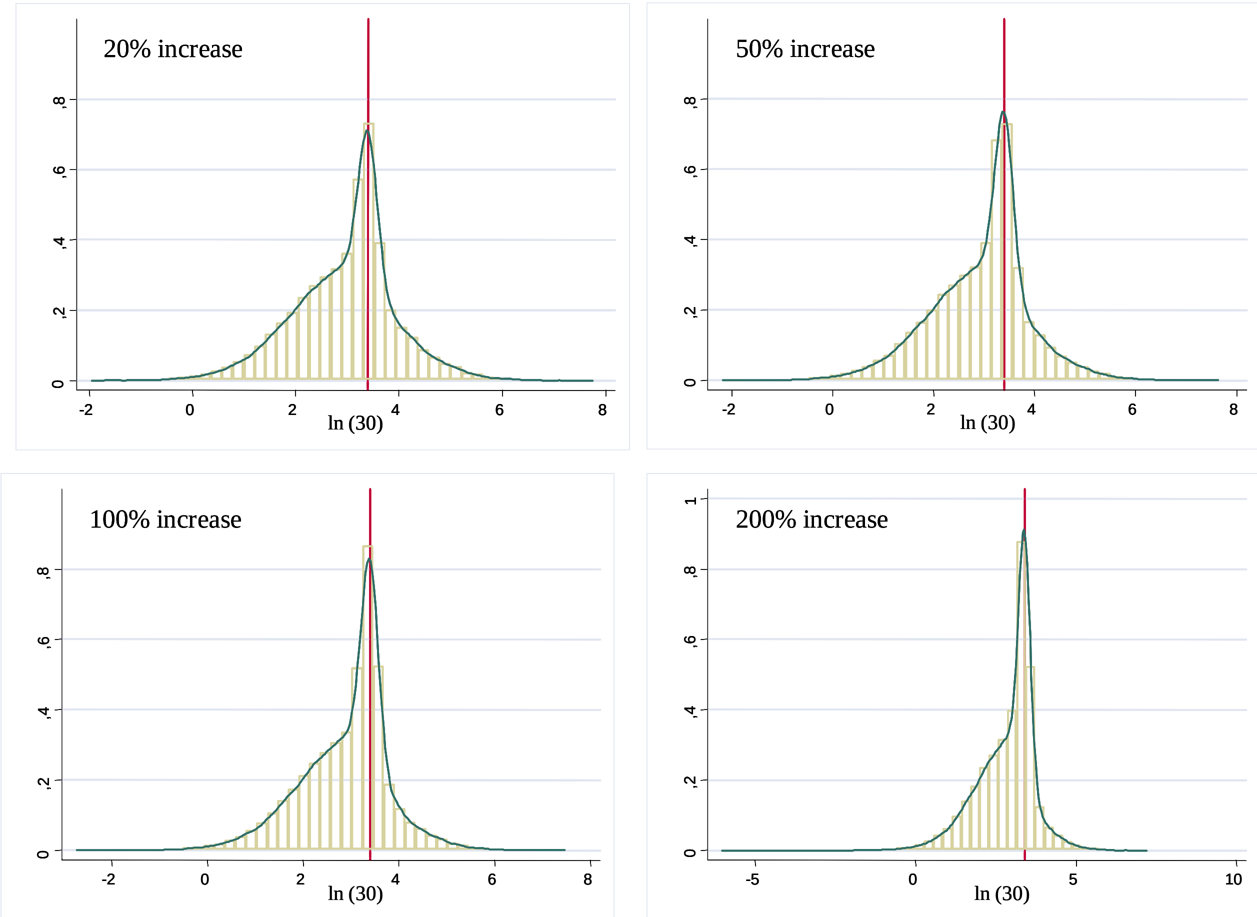

Figure 7: Histograms of cross-sectional distributions of simulated log-demands under simultaneous increases in p_1 and p_2. They are obtained by simulating 464260 observations under a 20%, 50%, 100% and 200% increase in both prices. Then, observations are averaged across individuals.

conclusions are very similar to the previous one, so I do not show the histograms. The more highlighting difference is with respect to the distributions corresponding to a change in p_1. They also show an increasing proportion of households consuming in the first block. However, this proportion is lower in comparison with figure 7. Thus, a change in p_1 reduces demands of people consuming at block two but the decrease is lower than under a rise in p_2 or in the two prices. These behaviours are coherent with the magnitude of the price-elasticities.

I complete the analysis of the effects of a rise in the two prices with table 16. It shows the predicted probabilities after the shock realizes for each one of the simultaneous change in the prices. As prices are higher, the probability of being at block two is lower and the probability of being at block one is bigger. Although the reduction is not very high, this table, jointly with figure 7 and table 15, indicates there are possibilities of reducing household demands by increasing prices.

Another possible policy analysis is to use the model to predict the effects of a change in the tariff. That is, what are the effects on the percentage of people at each block if the number of blocks and the marginal prices change.

The tariff of Town is different since 2005. It has four blocks with new marginal prices (see table 1) but the fixed cost disappears. The database has consumptions in the first quarter of 2005. So, I can exploit these data to test the ability of the model to predict the effects of the new tariff.

The way I proceed is as follows. First, I construct the microeconomic model for the new tariff and derive the demand function (now, it has seven branches). I develop the empirical model in the same way than for the model with a two-block tariff. Imposing the same assumptions on β_i and ε_{it}, I can derive the probabilities of being at each block before and after the shock[8]. Table 17 compares those probabilities with the percentages from data.

The model predicts well the direction of the change: the block with more individuals is the first, the second block with more people is the two and so on. The percentages differ a little from the observed ones. But this can be due to the fact I only have available one quarter of data with the new tariff. It is possible that some people are not aware of the new tariff until the end of the quarter, when the bill arrives to their houses. And it is also possible that other households know the change but they can not adjust their demands completely from one quarter to another due to the existence of habits in the water consumption.

[8] In appendix 4, I develop the model. Note that the demand in the first block will be some different because the fixed cost is not in the new tariff.

8 Conclusions and further research

This paper presents a methodology to analyse demands when the consumer faces an increasing block tariff and when only data on individual consumptions are available. I apply this methodology to analyse the individual water demand of a Spanish town. There, households face an increasing tariff with two blocks. I construct a model that explains the observed demands. The nonlinear price implies a kinked budget constraint and, consequently, water demand is also nonlinear.

The only information available in the database is about household consumptions. Without income data, it is not possible to estimate demands. The empirical strategy consists of assuming the household income is constant over the period covered by the database. The income is treated as an unobservable individual effect that follows a normal distribution. The individual effect determines the block and the optimal quantity of water consumed by the household.

However, observed demands are not the optimization choices. There is a time-varying term, unobservable to the econometrician and to the household, that makes the observed demands are not the utility-maximizing locations. It allows reproducing the fact that almost nobody is observed at the kink.

This economic setup gives rise to a random effects model with a nonlinear individual effect.

The data analysis shows people consume below the kink in spite of a fixed cost in the tariff. This cost implies that a person has to pay the cost of thirty units although he consumes less than thirty. To account for this fact, I include an opportunity cost (c) in the budget constraint. If agents demand below thirty, this implies they have into account other factors, not in the tariff, when they decide their optimal water demand. With the opportunity cost, the relevant marginal price for an agent consuming at the first block is c instead of zero. The marginal price for those consuming at block two is $c + p_2$. The estimation of the opportunity cost is 0.583 euros/cubic meter and it is significant.

To obtain the estimations, I use a likelihood-based indirect inference method that uses as auxiliary model an approximation of the structural model.

The validity analysis shows the model fit is good. So, the introduction of the opportunity cost can explain the observed water demands. This cost will include several aspects such as the ecological awareness of the agent or the cost of controlling if he has already reached the kink.

Using the estimated parameters, I obtain the cross-sectional distribution of the price-elasticity. Its mean is -0.12. This is evidence water demand is quite inelastic. However, if I compute the mean elasticity by distinguishing people consuming at block one or two, conclusions change. The price-elasticity of those that consume above the kink is -0.45 while for those at the first block is -0.01. So, in presence of a fixed cost in the tariff, people demanding above thirty cubic meters are more responsiveness to a change in the two marginal prices than people demanding below.

The price-elasticity I obtain is similar to what Martinez-Espiñeira (2002) calculates. He uses data at the municipal level from several cities of the same Spanish region as Town. Unlike him, I have household data and I can construct a microeconomic model that explains in detail the behaviour of the households, specially in the presence of the fixed cost.

In addition, my results agree with those obtained in the literature: water demand is inelastic but the elasticity increases as we move to higher blocks.

Finally, I use the model to make some policy analysis. As both marginal prices increase, the distribution of the cross-sectional demands accumulates on the left of the kink. However, in spite of the prices increase a lot there are always some individuals consuming at the second block. This indicates they have water necessities that force them to demand above thirty cubic meters independently of how much prices increase.

The second policy analysis I consider is to know the effects of a change in the tariff. In 2005, a new tariff applies in Town. The database includes consumptions in the first quarter of 2005. Therefore, I exploit this information to test how the model predicts the effects of the new tariff. Predictions indicates correctly the direction of the change but the percentages of people at each block differ from what is observed. However, this can be due to some consumers may not still be aware of the new tariff. It also can be explained by the fact that other households can not adjust their demands completely from one quarter to another due to the existence of habits in the water consumption.

Although the model explains the household behaviour, there are some considerations to have into account in future work. In the paper, the seasonality is included in the error term. An alternative would consist of incorporating seasonality in the model, for example, by introducing four different coefficients (ν_1, ν_2, ν_3, ν_4) in the utility function, one for each quarter.

The error has persistence but in the model, shocks are considered independent. It is possible to include the persistency allowing autocorrelated shocks.

On the other hand, ε_{it} could be incorporated to the water demand like an anticipated shock because it is reasonable to think individuals do not wrong systematically when they choose their consumptions.

To conclude, the methodology developed in this paper is not water specific. It can be used to analyse individual demands of whatever good with an increasing-block pricing schedule. Thus, for example, it can be applied to other utilities such as gas, electricity, telephone. And only data on individual consumptions are needed. In the case of utilities, the advantage is those data do not involve surveys. They are simply collected from household meters.

A Appendix 1: Auxiliary model

The individual likelihood in the auxiliary model is[9]:

$$\begin{aligned} f(y_{i1}, y_{i2}, ..., y_{iT_i}) &= \int_{-\infty}^{\log 30+\log c-\log\nu} f(y_{i1}, y_{i2}, ..., y_{iT_i}/\beta_i) f(\beta_i) d\beta_i \\ &+ \int_{\log 30+\log c-\log\nu}^{\log 30+\log(c+p_2)-\log\nu} f(y_{i1}, y_{i2}, ..., y_{iT_i}/\beta_i) f(\beta_i) d\beta_i \\ &+ \int_{\log 30+\log(c+p_2)-\log\nu}^{+\infty} f(y_{i1}, y_{i2}, ..., y_{iT_i}/\beta_i) f(\beta_i) d\beta_i \end{aligned}$$

1. *First integral*: $a_i = \beta_i - \log c + \log \nu$.

$$\int_{-\infty}^{\log 30+\log c-\log\nu} f(y_{i1}, y_{i2}, ..., y_{iT_i}/\beta_i) f(\beta_i) d\beta_i = \int_{-\infty}^{\log 30+\log c-\log\nu} f(y_i/\beta_i) f(\beta_i) d\beta_i$$

If we take into account that $f(y_i/\beta_i, \overline{y}_i) = f(y_i/\overline{y}_i)$ because $\overline{y}_i$ is a sufficient statistic for β_i[10], therefore, it is true that:

$$f(y_i/\beta_i, \overline{y}_i) = \frac{f(y_i, \overline{y}_i/\beta_i)}{f(\overline{y}_i/\beta_i)} = \frac{f(y_i/\beta_i)}{f(\overline{y}_i/\beta_i)} = f(y_i/\overline{y}_i)$$

Using the previous expression, the integral can be rewritten in the following way:

$$\int_{-\infty}^{\log 30+\log c-\log\nu} f(y_i/\beta_i) f(\beta_i) d\beta_i = \int_{-\infty}^{\log 30+\log c-\log\nu} f(y_i/\overline{y}_i) f(\overline{y}_i/\beta_i) f(\beta_i) d\beta_i =$$

$$= f(y_i/\overline{y}_i) \int_{-\infty}^{\log 30+\log c-\log\nu} f(\overline{y}_i/\beta_i) f(\beta_i) d\beta_i = f(y_i/\overline{y}_i) \int_{-\infty}^{\log 30+\log c-\log\nu} f(\beta_i/\overline{y}_i) f(\overline{y}_i) d\beta_i =$$

$$= f(y_i/\overline{y}_i) f(\overline{y}_i) \int_{-\infty}^{\log 30+\log c-\log\nu} f(\beta_i/\overline{y}_i) d\beta_i = f(y_i) p(\beta_i \leq \log 30 + \log c - \log \nu/\overline{y}_i)$$

a) $f(y_i)$ is simply the normal multivariate density function:

$$\begin{aligned} a_i &\sim N\left(\mu_\beta + \mu,\ \sigma_\beta^2\right), \quad \mu \equiv \log\nu - \log c \\ \varepsilon_{it} &\sim iid\ N\left(0,\ \sigma_\varepsilon^2\right) \\ y_{it} &= a_i + \varepsilon_{it} \sim N\left(\mu_\beta + \mu,\ \sigma_\beta^2 + \sigma_\varepsilon^2\right) \end{aligned}$$

Consequently, the y_i vector follows a normal multivariate distribution:

$$y_i = \begin{pmatrix} y_{i1} \\ . \\ . \\ y_{iT_i} \end{pmatrix} \sim N_{multivariante} \left[\underbrace{\begin{pmatrix} \mu_\beta + \mu \\ . \\ . \\ \mu_\beta + \mu \end{pmatrix}}_{\overline{\mu}}, \underbrace{\begin{pmatrix} \sigma_\beta^2 + \sigma_\varepsilon^2 & \sigma_\beta^2 & . & \sigma_\beta^2 \\ \sigma_\beta^2 & . & . & \sigma_\beta^2 \\ . & . & . & . \\ \sigma_\beta^2 & . & \sigma_\beta^2 & \sigma_\beta^2 + \sigma_\varepsilon^2 \end{pmatrix}}_{\Omega} \right]$$

[9] For notational simplification, since now, $\log y_{it}$ will be simply y_{it} and $\log a_i$ will be a_i.

[10] Proof in Arellano, M. (2003), "Panel Data Econometrics", Oxford University Press, page 25.

where $\overline{\mu}$ is a $T_i \times 1$ mean vector and Ω is a $T_i \times T_i$ variances-covariances matrix. The y_i density function is:

$$f(y_i) = (2\pi)^{-(T_i/2)}(\det\Omega)^{-(1/2)}\exp\left[-\frac{1}{2}(y_i-\overline{\mu})'\Omega^{-1}(y_i-\overline{\mu})\right]$$

b) $p(\beta_i \leq \log 30 + \log c - \log\nu/\overline{y}_i)$:

Taking temporal averages in $y_{it} = a_i + \varepsilon_{it}$, I obtain $\overline{y}_i = \overline{a_i} + \overline{\varepsilon_i} = a_i + \overline{\varepsilon_i}$. Therefore, $\overline{y}_i \sim N\left(\mu_\beta + \mu, \sigma^2_\beta + \frac{\sigma^2_\varepsilon}{T_i}\right)$ and it is possible to define the joint distribution of $\overline{y}_i$ and β_i :

$$\begin{pmatrix} \overline{y}_i \\ \beta_i \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_\beta + \mu \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma^2_\beta + \frac{\sigma^2_\varepsilon}{T_i} & \sigma^2_\beta \\ \sigma^2_\beta & \sigma^2_\beta \end{pmatrix}\right]$$

So, $\beta_i/\overline{y}_i \sim N\left[(1-\gamma)\mu_\beta + \gamma(\overline{y}_i - \mu),\ \sigma^2_\beta(1-\gamma)\right]$, with $\gamma = \frac{\sigma^2_\beta}{\sigma^2_\beta + \frac{\sigma^2_\varepsilon}{T_i}}$.

And, $p(\beta_i \leq \log 30 - \mu/\overline{y}_i) = \Phi\left[\frac{\log 30 - \mu - \gamma(\overline{y}_i - \mu) - (1-\gamma)\mu_\beta}{\sigma_\beta\sqrt{(1-\gamma)}}\right]$.

2. *Second integral*: $a_i = \log 30$.

$$\int_{\log 30+\log c-\log\nu}^{\log 30+\log(c+p_2)-\log\nu} f(y_{i1}, y_{i2}, ..., y_{iT_i}/\beta_i)f(\beta_i)d\beta_i = \int_{\log 30+\log c-\log\nu}^{\log 30+\log(c+p_2)-\log\nu} \Pi_{t=1}^{T_i} f(y_{it}/\beta_i)f(\beta_i)d\beta_i =$$

$$= \int_{\log 30+\log c-\log\nu}^{\log 30+\log(c+p_2)-\log\nu} \left[\Pi_{t=1}^{T_i}\frac{1}{\sigma_\varepsilon}\phi\left(\frac{y_{it}-\log 30}{\sigma_\varepsilon}\right)\right] p_2 d\beta_i =$$

$$= \left[\Pi_{t=1}^{T_i}\frac{1}{\sigma_\varepsilon}\phi\left(\frac{y_{it}-\log 30}{\sigma_\varepsilon}\right)\right] p_2\left[\log(c+p_2) - \log(c)\right]$$

3. *Third integral*: $a_i = \beta_i - \log(c+p_2) + \log\nu$.

$$\int_{\log 30+\log(c+p_2)-\log\nu}^{+\infty} f(y_{i1}, y_{i2}, ..., y_{iT_i}/\beta_i)f(\beta_i)d\beta_i = \int_{\log 30+\log(c+p_2)-\log\nu}^{+\infty} f(y_i/\beta_i)f(\beta_i)d\beta_i$$

Following the same steps as in the first integral:

$$\int_{\log 30+\log(c+p_2)-\log\nu}^{+\infty} f(y_i/\beta_i)f(\beta_i)d\beta_i = f(y_i/\overline{y}_i)f(\overline{y}_i)\int_{\log 30+\log(c+p_2)-\log\nu}^{+\infty} f(\beta_i/\overline{y}_i)d\beta_i =$$

$$= f(y_i)\ p[\beta_i \geq \log 30 + \log(c+p_2) - \log\nu/\overline{y}_i]$$

a) $f(y_i)$ is the normal multivariate density function:

$$\begin{aligned} a_i &\sim N\left(\mu_\beta + \mu',\ \sigma^2_\beta\right), \quad \mu' \equiv \log\nu - \log(c+p_2) \\ \varepsilon_{it} &\sim iid\ N\left(0,\ \sigma^2_\varepsilon\right) \\ y_{it} &= a_i + \varepsilon_{it} \sim N\left(\mu_\beta + \mu',\ \sigma^2_\beta + \sigma^2_\varepsilon\right) \end{aligned}$$

Consequently, the y_i vector follows a normal multivariate distribution:

$$y_i = \begin{pmatrix} y_{i1} \\ . \\ . \\ y_{iT_i} \end{pmatrix} \sim N_{multivariante} \left[\underbrace{\begin{pmatrix} \mu_\beta + \mu' \\ . \\ . \\ \mu_\beta + \mu' \end{pmatrix}}_{\overline{\mu'}}, \underbrace{\begin{pmatrix} \sigma_\beta^2 + \sigma_\varepsilon^2 & \sigma_\beta^2 & . & \sigma_\beta^2 \\ \sigma_\beta^2 & . & . & \sigma_\beta^2 \\ . & . & . & . \\ \sigma_\beta^2 & . & \sigma_\beta^2 & \sigma_\beta^2 + \sigma_\varepsilon^2 \end{pmatrix}}_{\Omega} \right]$$

where $\overline{\mu'}$ is a $T_i \times 1$ mean vector and Ω is a $T_i \times T_i$ variances-covariances matrix. The y_i density function is:

$$f(y_i) = (2\pi)^{-(T_i/2)} (\det \Omega)^{-(1/2)} \exp \left[-\frac{1}{2} (y_i - \overline{\mu'})' \Omega^{-1} (y_i - \overline{\mu'}) \right]$$

b) $p[\beta_i \geq \log 30 + \log(c + p_2) - \log \nu / \overline{y}_i]$:

In this case $\overline{y}_i \sim N\left(\mu_\beta + \mu', \sigma_\beta^2 + \frac{\sigma_\varepsilon^2}{T_i}\right)$. The joint distribution of $\overline{y}_i$ and β_i is, therefore:

$$\begin{pmatrix} \overline{y}_i \\ \beta_i \end{pmatrix} \sim N \left[\begin{pmatrix} \mu_\beta + \mu' \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\beta^2 + \frac{\sigma_\varepsilon^2}{T_i} & \sigma_\beta^2 \\ \sigma_\beta^2 & \sigma_\beta^2 \end{pmatrix} \right]$$

So, $\beta_i / \overline{y}_i \sim N\left[(1-\gamma)\mu_\beta + \gamma(\overline{y}_i - \mu'),\ \sigma_\beta^2 (1-\gamma)\right]$, with $\gamma = \frac{\sigma_\beta^2}{\sigma_\beta^2 + \frac{\sigma_\varepsilon^2}{T_i}}$.

With this distribution, the conditional probability is:

$p[\beta_i \geq \log 30 + \log(c+p_2) - \log \nu / \overline{y}_i] = p(\beta_i \geq \log 30 - \mu' / \overline{y}_i) = 1 - p(\beta_i < \log 30 - \mu' / \overline{y}_i) = 1 - \Phi\left[\frac{\log 30 - \mu' - \gamma(\overline{y}_i - \mu') - (1-\gamma)\mu_\beta}{\sigma_\beta \sqrt{(1-\gamma)}}\right]$.

Therefore, the individual likelihood is:

$$\begin{aligned} f(y_i) = {} & (2\pi)^{-(T_i/2)} (\det \Omega)^{-(1/2)} \exp \left[-\frac{1}{2} (y_i - \overline{\mu})' \Omega^{-1} (y_i - \overline{\mu}) \right] \\ & \times \Phi \left[\frac{\log 30 - \mu - \gamma(\overline{y}_i - \mu) - (1-\gamma)\mu_\beta}{\sigma_\beta \sqrt{(1-\gamma)}} \right] \\ & + \left[\Pi_{t=1}^{T_i} \frac{1}{\sigma_\varepsilon} \phi \left(\frac{y_{it} - \log 30}{\sigma_\varepsilon} \right) \right] p_2 \left[\log(c + p_2) - \log(c) \right] \\ & + (2\pi)^{-(T_i/2)} (\det \Omega)^{-(1/2)} \exp \left[-\frac{1}{2} (y_i - \overline{\mu'})' \Omega^{-1} (y_i - \overline{\mu'}) \right] \\ & \times \Phi \left[\frac{-\log 30 + \mu' + \gamma(\overline{y}_i - \mu') + (1-\gamma)\mu_\beta}{\sigma_\beta \sqrt{(1-\gamma)}} \right] \end{aligned}$$

The log-likelihood is: $L^{aux}(y_1, y_2, ..., y_N) = \sum_{i=1}^N \log f(y_i)$.

B Appendix 2: Standard errors

Let be the log-likelihood of the auxiliary model

$$L^{aux}\left[\theta;\log y(\alpha)\right]=\sum_{i=1}^{N}l_i(\theta)$$

where N is the number of individuals, θ and α are the vectors of auxiliary and structural parameters respectively and $\log y(\alpha)$ is the vector of data.

$$\hat{\theta}=\arg\max_{\theta} L^{aux}\left[\theta;\log y(\alpha)\right]$$

where plim $\hat{\theta}=\theta_0$. The robust estimation of the asymptotic variance is

$$\widehat{Var}(\hat{\theta})=\frac{1}{N}\hat{H}^{-1}\hat{W}\hat{H}^{-1}$$

where $\hat{H}$ and $\hat{W}$ are consistent estimations of

$$H=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2 l_i(\theta_0)}{\partial\theta\partial\theta'}$$

$$W=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial l_i(\theta_0)}{\partial\theta}\frac{\partial l_i(\theta_0)}{\partial\theta'}\right)$$

$L^{aux}\left[\theta;\log y(\alpha)\right]$ is a pseudo-likelihood, so, in general, the information equality does not hold and $W\neq H$.

The variance of $\widehat{\alpha}=\arg\max_{\alpha} L^{aux}\left[\widetilde{\theta}(\alpha);\log y\right]$ is:

$$\widehat{Var}(\widehat{\alpha})=\frac{1}{N}\left(\hat{D}'\hat{H}\hat{D}\right)^{-1}\hat{D}'\hat{W}\hat{D}\left(\hat{D}'\hat{H}\hat{D}\right)^{-1}$$

where $\hat{D}$ is the matrix of numerical partial derivatives evaluated in $\widehat{\alpha}$, $\hat{D}=\frac{\partial\widetilde{\theta}(\widehat{\alpha})}{\partial\alpha'}$.

When the number of parameters in α and θ is the same and $\hat{D}$ is invertible:

$$\widehat{Var}(\widehat{\alpha})=\frac{1}{N}\hat{D}^{-1}\widehat{Var}(\hat{\theta})\hat{D}^{-1\prime}$$

To calculate the standard errors of the structural model of section 5, I use this last expression because the auxiliary and the structural models have four parameters and $\hat{D}$ is a 4×4 invertible matrix.

As I reparametrize c and ν, $c=e^{c'}$ and $\nu=\frac{e^{\nu'}}{1+e^{\nu'}}$, the vector of structural parameters is

$$\widehat{\alpha}=\begin{pmatrix}\widehat{\sigma}^2_{\beta}\\ \widehat{\sigma}^2_{\varepsilon}\\ \widehat{c}'\\ \widehat{\nu}'\end{pmatrix}$$

So, once I obtain $\widehat{Var}(\widehat{\alpha})$, I applied the delta method to recover the standard errors of c and ν.

C Appendix 3: Probabilities after ε_{it}

The observed data are the result of combining the optimal decision of the agent with the error: $\log y_{it} = \log a_i + \varepsilon_{it}$.

Therefore,

$$\log y_{it} = \begin{cases} \log \nu - \log c + \log(e^{\beta_i} - FC) + \varepsilon_{it} \\ \quad if \quad \beta_i < \log\left(\frac{30c}{\nu} + FC\right) \\ \log 30 + \varepsilon_{it} \\ \quad if \;\; \log\left(\frac{30c}{\nu} + FC\right) < \beta_i < \log\left(\frac{30(c+p_2)}{\nu} + FC - 30p_2\right) \\ \log \nu - \log(c+p_2) + \log(e^{\beta_i} - FC + 30p_2) + \varepsilon_{it} \\ \quad if \quad \beta_i > \log\left(\frac{30(c+p_2)}{\nu} + FC - 30p_2\right) \end{cases}$$

Obtaining the unconditional probability of $\log y_{it}$ requires, previously, to define the cumulated distribution function of $\log y_{it}$ conditional on ε_{it}. This distribution has two branches and an accumulation point at $\log 30 + \varepsilon_{it}$.

Let r be any value from $\log y_{it}$ and $R = e^r$. I also define $\overline{r} \equiv \log 30 + \varepsilon_{it}$. I determine $p(\log y_{it} \leq r/\varepsilon_{it})$ as a function of r and ε_{it} for each one of the next possible cases:

1. $r < \overline{r}$. In that case, $\log y_{it} = \log \nu - \log c + \log(e^{\beta_i} - FC) + \varepsilon_{it}$.

   Therefore, $p(\log y_{it} \leq r/\varepsilon_{it}) = p\left[\log \nu - \log c + \log(e^{\beta_i} - FC) + \varepsilon_{it} \leq r/\varepsilon_{it}\right] =$
   $= p\left[\beta_i \leq \log\left(\frac{Rc}{\nu e^{\varepsilon_{it}}} + FC\right)\right] = \Phi\left[\frac{\log\left(\frac{Rc}{\nu e^{\varepsilon_{it}}} + FC\right) - \mu_\beta}{\sigma_\beta}\right] \equiv F_1(\log y_{it}/\varepsilon_{it})$.

2. $r = \overline{r}$. In that case, $\log y_{it} = \log 30 + \varepsilon_{it}$.

   Therefore, $p(\log y_{it} \leq \overline{r}/\varepsilon_{it}) = p(\log y_{it} < \overline{r}/\varepsilon_{it}) + p(\log y_{it} = \overline{r}/\varepsilon_{it}) = p\left[\beta_i \leq \log\left(\frac{\overline{R}c}{\nu e^{\varepsilon_{it}}} + FC\right)\right] +$
   $p\left[\log\left(\frac{30c}{\nu} + FC\right) < \beta_i < \log\left(\frac{30(c+p_2)}{\nu} + FC - 30p_2\right)\right] =$
   $\Phi\left[\frac{\log\left(\frac{\overline{R}c}{\nu e^{\varepsilon_{it}}} + FC\right) - \mu_\beta}{\sigma_\beta}\right] + \Phi\left[\frac{\log\left(\frac{30(c+p_2)}{\nu} + FC\right) - \mu_\beta}{\sigma_\beta}\right] - \Phi\left[\frac{\log\left(\frac{30c}{\nu} + FC\right) - \mu_\beta}{\sigma_\beta}\right] \equiv F_2(\log y_{it}/\varepsilon_{it})$.

3. $r > \overline{r}$. In that case, $\log y_{it} = \log \nu - \log(c+p_2) + \log(e^{\beta_i} - FC + 30p_2) + \varepsilon_{it}$.

   Therefore, $p(\log y_{it} \leq r/\varepsilon_{it}) = p(\log y_{it} < \overline{r}/\varepsilon_{it}) + p(\log y_{it} = \overline{r}/\varepsilon_{it}) +$
   $p(\overline{r} < \log y_{it} \leq r/\varepsilon_{it}) = \Phi\left[\frac{\log\left(\frac{\overline{R}c}{\nu e^{\varepsilon_{it}}} + FC\right) - \mu_\beta}{\sigma_\beta}\right] + \Phi\left[\frac{\log\left(\frac{30(c+p_2)}{\nu} + FC\right) - \mu_\beta}{\sigma_\beta}\right] - \Phi\left[\frac{\log\left(\frac{30c}{\nu} + FC\right) - \mu_\beta}{\sigma_\beta}\right] +$
   $\Phi\left[\frac{\log\left(\frac{R(c+p_2)}{\nu e^{\varepsilon_{it}}} + FC - 30p_2\right) - \mu_\beta}{\sigma_\beta}\right] - \Phi\left[\frac{\log\left(\frac{\overline{R}(c+p_2)}{\nu e^{\varepsilon_{it}}} + FC - 30p_2\right) - \mu_\beta}{\sigma_\beta}\right] \equiv F_3(\log y_{it}/\varepsilon_{it})$.

So,

$$F(\log y_{it}/\varepsilon_{it}) = \begin{cases} F_1(\log y_{it}/\varepsilon_{it}) & if\ r < \overline{r} \\ F_2(\log y_{it}/\varepsilon_{it}) & if\ r = \overline{r} \\ F_3(\log y_{it}/\varepsilon_{it}) & if\ r > \overline{r} \end{cases}$$

The unconditional probability is the average of the M conditional probabilities calculated by simulating values of ε_{it}:

$$p(\log y_{it} \leq r) = \tfrac{1}{M}\sum_{j=1}^{M} p(\log y_{it} \leq r/\varepsilon_{jt}).$$

I obtain 100000 draws from the normal distribution $N(0, \widehat{\sigma}_{\varepsilon}^2)$, with $\widehat{\sigma}_{\varepsilon}^2$ equal to the value from table 10. I calculate the conditional probability for $r = \log 30$ using the previous expressions, depending on the value of ε_{jt} and substituying the parameters by their estimations from table 10.

Finally, $p(block\ 1) = p(\log y_{it} \leq \log 30)$ and $p(block\ 2) = 1 - p(block\ 1)$.

D Appendix 4: The model for the tariff in 2005_Q1

Utility maximization problem with the new tariff:

$$\begin{aligned} \underset{\{rg,\, a\}}{Max} \quad u(rg, a) &= (rg)^{1-\nu} a^{\nu} \\ st. \qquad m &= rg + ca + p_1 a && if\ a \leq 25 \\ m &= rg + 25p_1 + ca + (a-25)p_2 && if\ 25 < a \leq 45 \\ m &= rg + 25p_1 + 20p_2 + ca + (a-45)p_3 && if\ 45 < a \leq 90 \\ m &= rg + 25p_1 + 20p_2 + 45p_3 + ca + (a-90)p_4 && if\ a > 90 \end{aligned}$$

Demand at each block:

1. Block 1: $a_1(m) = \frac{\nu}{c+p_1} m.$
2. Block 2: $a_2(m) = \frac{\nu}{c+p_2}\left(m - 25p_1 + 25p_2\right).$
3. Block 3: $a_3(m) = \frac{\nu}{c+p_3}\left(m - 25p_1 - 20p_2 + 45p_3\right).$
4. Block 4: $a_4(m) = \frac{\nu}{c+p_4}\left(m - 25p_1 - 20p_2 - 45p_3 + 90p_4\right).$

Linearizing the demand and assuming $m_i = e^{\beta_i}$, with $\beta_i \sim N\left(\mu_\beta, \sigma_\beta^2\right)$, the demand function is:

$$\log a_i = \begin{cases} \log\nu - \log(c+p_1) + \beta_i \\ if \beta_i < \log\left(\frac{25(c+p_1)}{\nu}\right) \\ \\ \log 25 \\ if \log\left(\frac{25(c+p_1)}{\nu}\right) < \beta_i < \log\left(\frac{25(c+p_2)}{\nu} + 25p_1 - 25p_2\right) \\ \\ \log\nu - \log(c+p_2) + \log\left(e^{\beta_i} - 25p_1 + 25p_2\right) \\ if \log\left(\frac{25(c+p_2)}{\nu} + 25p_1 - 25p_2\right) < \beta_i < \log\left(\frac{45(c+p_2)}{\nu} + 25p_1 - 25p_2\right) \\ \\ \log 45 \\ if \log\left(\frac{45(c+p_2)}{\nu} + 25p_1 - 25p_2\right) < \beta_i < \log\left(\frac{45(c+p_3)}{\nu} + 25p_1 + 20p_2 - 45p_3\right) \\ \\ \log\nu - \log(c+p_3) + \log\left(e^{\beta_i} - 25p_1 - 20p_2 + 45p_3\right) \\ if \log\left(\frac{45(c+p_3)}{\nu} + 25p_1 + 20p_2 - 45p_3\right) < \beta_i < \log\left(\frac{90(c+p_3)}{\nu} + 25p_1 + 20p_2 - 45p_3\right) \\ \\ \log 90 \\ if \log\left(\frac{90(c+p_3)}{\nu} + 25p_1 + 20p_2 - 45p_3\right) < \beta_i < \log\left(\frac{90(c+p_4)}{\nu} + 25p_1 + 20p_2 + 45p_3 - 90p_4\right) \\ \\ \log\nu - \log(c+p_4) + \log\left(e^{\beta_i} - 25p_1 - 20p_2 - 45p_3 + 90p_4\right) \\ if \beta_i > \log\left(\frac{90(c+p_4)}{\nu} + 25p_1 + 20p_2 + 45p_3 - 90p_4\right) \end{cases}$$

The probability of each branch is:

1. $p_1 \equiv p\left[\beta_i < \log\left(\frac{25(c+p_1)}{\nu}\right)\right] = \Phi\left[\frac{\log 25 + \log(c+p_1) - \log\nu - \mu_\beta}{\sigma_\beta}\right]$.
2. $p_{25} \equiv p\left[\log\left(\frac{25(c+p_1)}{\nu}\right) < \beta_i < \log\left(\frac{25(c+p_2)}{\nu} + 25p_1 - 25p_2\right)\right] = \Phi\left[\frac{\log\left(\frac{25(c+p_2)}{\nu} + 25p_1 - 25p_2\right) - \mu_\beta}{\sigma_\beta}\right] - \Phi\left[\frac{\log 25 + \log(c+p_1) - \log\nu - \mu_\beta}{\sigma_\beta}\right]$.
3. $p_2 \equiv p\left[\log\left(\frac{25(c+p_2)}{\nu} + 25p_1 - 25p_2\right) < \beta_i < \log\left(\frac{45(c+p_2)}{\nu} + 25p_1 - 25p_2\right)\right] = \Phi\left[\frac{\log\left(\frac{45(c+p_2)}{\nu} + 25p_1 - 25p_2\right) - \mu_\beta}{\sigma_\beta}\right] - \Phi\left[\frac{\log\left(\frac{25(c+p_2)}{\nu} + 25p_1 - 25p_2\right) - \mu_\beta}{\sigma_\beta}\right]$.
4. $p_{45} \equiv p\left[\log\left(\frac{45(c+p_2)}{\nu} + 25p_1 - 25p_2\right) < \beta_i < \log\left(\frac{45(c+p_3)}{\nu} + 25p_1 + 20p_2 - 45p_3\right)\right] = = \Phi\left[\frac{\log\left(\frac{45(c+p_3)}{\nu} + 25p_1 + 20p_2 - 45p_3\right) - \mu_\beta}{\sigma_\beta}\right] - \Phi\left[\frac{\log\left(\frac{45(c+p_2)}{\nu} + 25p_1 - 25p_2\right) - \mu_\beta}{\sigma_\beta}\right]$.
5. $p_3 \equiv p\left[\log\left(\frac{45(c+p_3)}{\nu} + 25p_1 + 20p_2 - 45p_3\right) < \beta_i < \log\left(\frac{90(c+p_3)}{\nu} + 25p_1 + 20p_2 - 45p_3\right)\right] =$

$$= \Phi\left[\frac{\log\left(\frac{90(c+p_3)}{\nu}+25p_1+20p_2-45p_3\right)-\mu_\beta}{\sigma_\beta}\right] - \Phi\left[\frac{\log\left(\frac{45(c+p_3)}{\nu}+25p_1+20p_2-45p_3\right)-\mu_\beta}{\sigma_\beta}\right].$$

6. $p_{90} \equiv p\left[\log\left(\frac{90(c+p_3)}{\nu}+25p_1+20p_2-45p_3\right) < \beta_i < \log\left(\frac{90(c+p_4)}{\nu}+25p_1+20p_2+45p_3-90p_4\right)\right] =$

$$= \Phi\left[\frac{\log\left(\frac{90(c+p_4)}{\nu}+25p_1+20p_2+45p_3-90p_4\right)-\mu_\beta}{\sigma_\beta}\right] - \Phi\left[\frac{\log\left(\frac{90(c+p_3)}{\nu}+25p_1+20p_2-45p_3\right)-\mu_\beta}{\sigma_\beta}\right].$$

7. $p_4 \equiv p\left[\beta_i > \log\left(\frac{90(c+p_4)}{\nu}+25p_1+20p_2+45p_3-90p_4\right)\right] =$ $1-\Phi\left[\frac{\log\left(\frac{90(c+p_4)}{\nu}+25p_1+20p_2+45p_3-90p_4\right)-\mu_\beta}{\sigma_\beta}\right].$

Observed data are equal to: $\log y_{it} = \log a_i + \varepsilon_{it}$, where $\varepsilon_{it} \sim iid\ N\left(0, \sigma_\varepsilon^2\right)$.

To obtain the unconditional probability of $\log y_{it}$, I follow the same process as in appendix 3. Let r be any value from $\log y_{it}$ and $R = e^r$. I also define $\overline{r_1} \equiv \log 25 + \varepsilon_{it}$, $\overline{r_2} \equiv \log 45 + \varepsilon_{it}$ and $\overline{r_3} \equiv \log 90 + \varepsilon_{it}$.

I determine $p(\log y_{it} \leq r/\varepsilon_{it})$ as a function of r and ε_{it} for each one of the next possible cases:

1. $r < \overline{r_1}$: $\log y_{it} = \log\nu - \log(c+p_1) + \beta_i + \varepsilon_{it}$.

   $p(\log y_{it} \leq r/\varepsilon_{it}) = p\left[\log\nu - \log(c+p_1) + \beta_i + \varepsilon_{it} \leq r/\varepsilon_{it}\right] = p\left[\beta_i \leq r - \varepsilon_{it} - \log\nu + \log(c+p_1)\right] =$ $\Phi\left[\frac{r-\varepsilon_{it}-\log\nu+\log(c+p_1)-\mu_\beta}{\sigma_\beta}\right] \equiv F_1(\log y_{it}/\varepsilon_{it})$.

2. $r = \overline{r_1}$: $\log y_{it} = \log 25 + \varepsilon_{it}$.

   $p(\log y_{it} \leq \overline{r_1}/\varepsilon_{it}) = p(\log y_{it} < \overline{r_1}/\varepsilon_{it}) + p(\log y_{it} = \overline{r_1}/\varepsilon_{it}) = \Phi\left[\frac{\overline{r_1}-\varepsilon_{it}-\log\nu+\log(c+p_1)-\mu_\beta}{\sigma_\beta}\right] +$ $p_{25} \equiv F_2(\log y_{it}/\varepsilon_{it})$.

3. $\overline{r_1} < r < \overline{r_2}$: $\log y_{it} = \log\nu - \log(c+p_2) + \log(e^{\beta_i} - 25p_1 + 25p_2) + \varepsilon_{it}$.

   $p(\log y_{it} \leq r/\varepsilon_{it}) = p(\log y_{it} \leq \overline{r_1}/\varepsilon_{it}) + p(\overline{r_1} < \log y_{it} \leq r/\varepsilon_{it}) = F_2(\log y_{it}/\varepsilon_{it}) +$ $p(\log y_{it} \leq r/\varepsilon_{it}) - p(\log y_{it} \leq \overline{r_1}/\varepsilon_{it}) = F_2(\log y_{it}/\varepsilon_{it}) + \Phi\left[\frac{\log\left(\frac{R(c+p_2)}{\nu e^{\varepsilon_{it}}}+25p_1-25p_2\right)-\mu_\beta}{\sigma_\beta}\right] -$ $\Phi\left[\frac{\log\left(\frac{\overline{R_1}(c+p_2)}{\nu e^{\varepsilon_{it}}}+25p_1-25p_2\right)-\mu_\beta}{\sigma_\beta}\right] \equiv F_3(\log y_{it}/\varepsilon_{it})$.

4. $r = \overline{r_2}$: $\log y_{it} = \log 45 + \varepsilon_{it}$.

   $p(\log y_{it} \leq \overline{r_2}/\varepsilon_{it}) = p(\log y_{it} \leq \overline{r_1}/\varepsilon_{it}) + p(\overline{r_1} < \log y_{it} < \overline{r_2}/\varepsilon_{it}) + p(\log y_{it} = \overline{r_2}/\varepsilon_{it}) =$ $F_2(\log y_{it}/\varepsilon_{it}) + \Phi\left[\frac{\log\left(\frac{\overline{R_2}(c+p_2)}{\nu e^{\varepsilon_{it}}}+25p_1-25p_2\right)-\mu_\beta}{\sigma_\beta}\right] - \Phi\left[\frac{\log\left(\frac{\overline{R_1}(c+p_2)}{\nu e^{\varepsilon_{it}}}+25p_1-25p_2\right)-\mu_\beta}{\sigma_\beta}\right] + p_{45} \equiv$ $F_4(\log y_{it}/\varepsilon_{it})$.

5. $\overline{r_2} < r < \overline{r_3}$: $\log y_{it} = \log \nu - \log(c+p_3) + \log(e^{\beta_i} - 25p_1 - 20p_2 + 45p_3) + \varepsilon_{it}$.

$p(\log y_{it} \leq r/\varepsilon_{it}) = p(\log y_{it} \leq \overline{r_2}/\varepsilon_{it}) + p(\overline{r_2} < \log y_{it} \leq r/\varepsilon_{it}) = F_4(\log y_{it}/\varepsilon_{it}) + p(\log y_{it} \leq r/\varepsilon_{it}) - p(\log y_{it} \leq \overline{r_2}/\varepsilon_{it}) = F_4(\log y_{it}/\varepsilon_{it}) + \Phi\left[\frac{\log\left(\frac{R(c+p_3)}{\nu e^{\varepsilon_{it}}}+25p_1+20p_2-45p_3\right)-\mu_\beta}{\sigma_\beta}\right] - \Phi\left[\frac{\log\left(\frac{\overline{R_2}(c+p_3)}{\nu e^{\varepsilon_{it}}}+25p_1+20p_2-45p_3\right)-\mu_\beta}{\sigma_\beta}\right] \equiv F_5(\log y_{it}/\varepsilon_{it})$.

6. $r = \overline{r_3}$: $\log y_{it} = \log 90 + \varepsilon_{it}$.

$p(\log y_{it} \leq \overline{r_3}/\varepsilon_{it}) = p(\log y_{it} \leq \overline{r_2}/\varepsilon_{it}) + p(\overline{r_2} < \log y_{it} < \overline{r_3}/\varepsilon_{it}) + p(\log y_{it} = \overline{r_3}/\varepsilon_{it}) = F_4(\log y_{it}/\varepsilon_{it}) + \Phi\left[\frac{\log\left(\frac{\overline{R_3}(c+p_3)}{\nu e^{\varepsilon_{it}}}+25p_1+20p_2-45p_3\right)-\mu_\beta}{\sigma_\beta}\right] - \Phi\left[\frac{\log\left(\frac{\overline{R_2}(c+p_3)}{\nu e^{\varepsilon_{it}}}+25p_1+20p_2-45p_3\right)-\mu_\beta}{\sigma_\beta}\right] + p_{90} \equiv F_6(\log y_{it}/\varepsilon_{it})$.

7. $r > \overline{r_3}$: $\log y_{it} = \log \nu - \log(c+p_4) + \log(e^{\beta_i} - 25p_1 - 20p_2 - 45p_3 + 90p_4) + \varepsilon_{it}$.

$p(\log y_{it} \leq r/\varepsilon_{it}) = p(\log y_{it} \leq \overline{r_3}/\varepsilon_{it}) + p(\overline{r_3} < \log y_{it} \leq r/\varepsilon_{it}) = F_6(\log y_{it}/\varepsilon_{it}) + p(\log y_{it} \leq r/\varepsilon_{it}) - p(\log y_{it} \leq \overline{r_3}/\varepsilon_{it}) = F_6(\log y_{it}/\varepsilon_{it}) + \Phi\left[\frac{\log\left(\frac{R(c+p_4)}{\nu e^{\varepsilon_{it}}}+25p_1+20p_2+45p_3-90p_4\right)-\mu_\beta}{\sigma_\beta}\right] - \Phi\left[\frac{\log\left(\frac{\overline{R_3}(c+p_4)}{\nu e^{\varepsilon_{it}}}+25p_1+20p_2+45p_3-90p_4\right)-\mu_\beta}{\sigma_\beta}\right] \equiv F_7(\log y_{it}/\varepsilon_{it})$.

So,

$$F(\log y_{it}/\varepsilon_{it}) = \begin{cases} F_1(\log y_{it}/\varepsilon_{it}) & if\ r < \overline{r_1} \\ F_2(\log y_{it}/\varepsilon_{it}) & if\ r = \overline{r_1} \\ F_3(\log y_{it}/\varepsilon_{it}) & if\ \overline{r_1} < r < \overline{r_2} \\ F_4(\log y_{it}/\varepsilon_{it}) & if\ r = \overline{r_2} \\ F_5(\log y_{it}/\varepsilon_{it}) & if\ \overline{r_2} < r < \overline{r_3} \\ F_6(\log y_{it}/\varepsilon_{it}) & if\ r = \overline{r_3} \\ F_7(\log y_{it}/\varepsilon_{it}) & if\ r > \overline{r_3} \end{cases}$$

The unconditional probability is the average of the M conditional probabilities calculated by simulating values of ε_{it}:

$p(\log y_{it} \leq r) = \frac{1}{M}\sum_{j=1}^{M} p(\log y_{it} \leq r/\varepsilon_{jt})$.

I obtain 100000 draws from the normal distribution $N(0, \widehat{\sigma}_\varepsilon^2)$, with $\widehat{\sigma}_\varepsilon^2$ equal to the value from table 10. I calculate the unconditional probability for $r = \log 25$, $r = \log 45$ and $r = \log 90$ using the previous expressions, depending on the value of ε_{jt} and substituying the parameters by their estimations from table 10. The probability of being at each block is:

1. $p(block\ 1) = p(\log y_{it} \leq \log 25)$.

2. $p(block\ 2) = p(\log 25 < \log y_{it} \leq \log 45) = p(\log y_{it} \leq \log 45) - p(\log y_{it} < \log 25)$.

3. $p(block\ 3) = p(\log 45 < \log y_{it} \leq \log 90) = p(\log y_{it} \leq \log 90) - p(\log y_{it} < \log 45)$.

4. $p(block\ 4) = p(\log y_{it} > \log 90) = 1 - p(\log y_{it} \leq \log 90)$.

References

[1] Burtless, G. and J. A. Hausman (1978), "The Effect of Taxation on Labor Supply: Evaluating the Gary Income Maintenance Experiment", *Journal of Political Economy*, vol. 86, no. 6, 1103-1130.

[2] Dalhuisen, J.M., R.J.G.M. Florax, H.L.F.M. de Groot and P. Nijkamp (2001), "Price and Income Elasticities of Residential Water Demand: Why Empirical Estimates Differ", Tinbergen Institute Discussion Paper 057/3.

[3] Gourieroux, C., A. Monfort and E. Renault (1993), "Indirect Inference", *Journal of Applied Econometrics* 8, S85-S118.

[4] Hanemann, W. M. (1984), "Discrete/Continuous Models of Consumer Demand", *Econometrica* vol. 52, no. 3, 541-561.

[5] Hausman, J. A. (1979), "The Econometrics of Labor Supply on Convex Budget Sets", *Economics Letters* 3,171-174.

[6] Hausman, J. A. (1985), "The Econometrics of Nonlinear Budget Sets", *Econometrica* vol. 53, no. 6, 1255-1282.

[7] Hewitt, J. A. and W. M. Hanemann (1995), "A Discrete/Continuos Choice Approach to Residential Water Demand under Block Rate Pricing", *Land Economics*, vol. 71, no. 2, 173-192.

[8] Keane, M. and A. Smith (2003), "Generalized Indirect Inference for Discrete Choice Models", Yale University, Working Paper.

[9] Martínez-Espiñeira, R. (2002), "Residential Water Demand in the Northwest of Spain", *Environmental and Resource Economics,* vol. 21, no.2, 161-187.

[10] Moffit, R. (1990), "The Econometrics of Kinked Budget Constraints", *Journal of Economic Perspectives*, vol. 4, no. 2, 119-139.

[11] Nieswiadomy M. L. and Molina D. J. (1989), "Comparing Residential Water Demand Estimates under Decreasing and Increasing Block Rates Using Household Data", *Land Economics*, vol. 65, no. 3, 280-289.

[12] Nordin J. A. (1976), "A Proposed Modification of Taylor's Demand Analysis: Comment", *The Bell Journal of Economics*, vol. 7, no. 2, 719-721.

[13] Olmstead, S.M., W.M. Hanemann and R.N. Stavins (2007), "Water Demand under Alternative Price Structures", NBER Working Paper 13573.

[14] Pudney S (1989), *Modelling the Individual Choice. The Econometrics of Corners, Kinks and Holes*, Oxford, Basil Blackwell.

[15] Taylor L. D. (1975), "The Demand for Electricity: A Survey", *The Bell Journal of Economics*, vol. 6, no. 1, 74-110.

TABLES:

Table 1: Tariffs ($euros/m^3$)

	Consumption	Sewer System	Total
2001-2002-2003-2004			
Below 30 m^3	0.29	0.11	0.40
Above 30 m^3	0.33	0.15	0.48
2005			
Below 25 m^3	0.15	0.11	0.26
From 26 to 45 m^3	0.40	0.20	0.60
From 46 to 90 m^3	0.65	0.30	0.95
Above 90 m^3	0.90	0.45	1.35

Table 2: Panel description for period 2001_Q3 - 2004_Q4

No. of quarters	Households	Percentage	Observations
1	175	3.98	175
2	151	3.44	302
3	79	1.80	237
4	74	1.68	296
5	50	1.14	250
6	92	2.09	552
7	46	1.05	322
8	132	3.00	1056
9	66	1.50	594
10	81	1.84	810
11	115	2.62	1265
12	32	0.73	384
13	25	0.57	325
14	3276	74.56	45864
Total	4394	100	52432

Table 3: Panel description (without zero-consumptions)

No. of quarters	Households	Percentage	Observations
1	220	5.16	220
2	179	4.20	358
3	116	2.72	348
4	110	2.58	440
5	90	2.11	450
6	125	2.93	750
7	107	2.51	749
8	167	3.91	1336
9	96	2.25	864
10	137	3.21	1370
11	182	4.27	2002
12	140	3.28	1680
13	279	6.54	3627
14	2318	54.34	32452
Total	4266	100	46646

Table 4: Percentage of observations at the blocks and the kink

Block 1 (*below* $30m^3$)	53.6
Kink ($30m^3$)	2.4
Block 2 (*above* $30m^3$)	44
Total	100
Interval: [25, 35] m^3	23.5
Interval: [27, 33] m^3	15.2
Interval: [29, 31] m^3	6.6

Note: 46646 observations.

Table 5: Statistics of cross-sectional distribution of observed log-demands

Mean	2.99
Standard deviation	1.02
Skewness	-1.09
Kurtosis	4.06
Median	3.24
Percentile 5%	0.69
Percentile 25%	2.59
Percentile 75%	3.65
Percentile 95%	4.20

Note: ln(30) = 3.40. 46426 observations and 4046 households

Table 6: Mobility among blocks

	Individuals	%
All quarters at block 1	1152	28.47
All quarters at kink	0	0
All quarters at block 2	780	19.28
Not all quarters at the same block	2114	52.25
Total	4046[a]	100

[a]It does not include households with only one observation.

Table 7: Estimation of the seasonal RE model

Variables	Coefficients	Standard error
d_{II}	0.02*	0.01
d_{III}	0.04*	0.01
d_{IV}	0.02*	0.01
constant	2.95*	0.02

Note: The dependent variable is log-consumption.

RE: Random Effects. 46646 observations.

*Significant at the 5% level.

Table 8: Variances from the RE model

	Variance	% over $V(\log m_{it}^3)$
σ_η^2	1.10	79.25
σ_ε^2	0.29	20.75
$V(\log m_{it}^3)$	1.39	100

Note: 46646 observations. RE: Random Effects.

Table 9: Estimation of the transitory error persistence

Variables	Coefficients	Robust standard error
$(\log m_{it-1}^3)_D$	0.42*	0.04
$(\log m_{it-2}^3)_D$	0.02	0.02
$(\log m_{it-3}^3)_D$	0.05*	0.02
$(\log m_{it-4}^3)_D$	0.08*	0.01
$(\log m_{it-5}^3)_D$	-0.05*	0.01

Note: 46646 observations. Estimations were obtained with the *xtabond* stata command.

*Significant at the 5% level.

Table 10: Estimates of structural parameters

Parameters	Estimation	Robust standard error
σ^2_β	1.406*	0.001
σ^2_ε	0.291*	0.000
c	0.583*	0.002
ν	0.004*	0.000

Note: Estimations obtained by indirect inference method with $M = 4$ and $\lambda = 0.05$. Standard errors calculated as shown in appendix 2. 46426 observations.
*Significant at the 1% level.

Table 11: Cross-sectional distribution of the price-elasticities by households consuming...

	at block 1	at block 2
Mean	-0.0120	-0.4504
Standard Deviation	0.0164	0.0000
Skewness	-7.6216	-0.0705
Kurtosis	106.90	1.9299
Median	-0.0073	-0.4504
Percentile 5%	-0.0354	-0.4505
Percentile 25%	-0.0132	-0.4505
Percentile 75%	-0.0046	-0.4503
Percentile 95%	-0.0033	-0.4503
Households	22952	10133

Note: 464260 simulated observations corresponding to 40460 households.

Table 12: Probabilities

	According data†	Model prediction (before shock ε_{it})‡	Model prediction (after shock ε_{it})‡
Block 1 (*below* $30m^3$)	0.54	0.56	0.63
Kink ($30m^3$)	0.02	0.18	0.00
Block 2 (*above* $30m^3$)	0.44	0.25	0.37

†46646 observations.
‡Predictions calculated using 100000 draws from the normal distribution $N(0, \widehat{\sigma}^2_\varepsilon)$.

Table 13: Comparison between the observed and the simulated cross-sectional distributions

	Observed	Simulated
Mean	2.99	3.00
Standard Deviation	1.02	1.01
Pearson coefficient	0.34	0.34
Skewness	-1.09	-0.26
Kurtosis	4.06	3.80
Median	3.24	3.16
Percentile 5%	0.69	1.24
Percentile 25%	2.59	2.38
Percentile 75%	3.65	3.55
Percentile 95%	4.20	4.60
Households	4046	40460

Note: log(30)=3.40

Table 14: Price variation

	p_1	FC	p_2
Initial prices	0.40	12.13	0.48
Δ20%	0.49	14.55	0.58
Δ50%	0.61	18.19	0.72
Δ100%	0.81	24.25	0.96
Δ200%	1.21	36.38	1.44

Note: $FC\ (FixedCost) = 30p_1$

Table 15: Cross-sectional distributions under different variations in p_1 and p_2

	Δ20%	Δ50%	Δ100%	Δ200%
Mean	2.98	2.95	2.92	2.87
Standard Deviation	0.99	0.96	0.93	0.90
Pearson coefficient	0.33	0.33	0.32	0.31
Skewness	-0.33	-0.43	-0.60	-0.91
Median	3.16	3.15	3.15	3.14
Percentile 5%	1.23	1.22	1.21	1.18
Percentile 25%	2.38	2.38	2.37	2.37
Percentile 75%	3.52	3.50	3.47	3.44
Percentile 95%	4.51	4.40	4.23	3.97

Note: 40460 households and 464260 observations.

Table 16: Probabilities under different price variations

	p_1	p_2	*Probability Block 1*	*Probability Block 2*
Δ20%	0.49	0.58	0.64	0.36
Δ50%	0.61	0.72	0.65	0.35
Δ100%	0.81	0.96	0.66	0.34
Δ200%	1.21	1.44	0.68	0.32

Note: Probabilities calculated using 100000 draws from the normal distribution $N(0, \widehat{\sigma}_{\varepsilon}^2)$.

Table 17: Probabilities in 2005_Q1

	According data†	Model prediction (before shock ε_{it})‡	Model prediction (after shock ε_{it})‡
Block 1 (*below* $25m^3$)	0.47	0.62	0.66
Kink 1 ($25m^3$)	0.02	0.10	0.00
Block 2 (*from* 26 *to* 45 m^3)	0.35	0.14	0.18
Kink 2 ($45m^3$)	0.01	0.04	0.00
Block 3 (*from* 46 *to* 90 m^3)	0.13	0.07	0.11
Kink 3 ($90m^3$)	0.00	0.01	0.00
Block 4 (*above* $90m^3$)	0.01	0.02	0.05

†3861 observations. ‡Predictions calculated using 100000 draws from the normal distribution $N(0, \widehat{\sigma}_{\varepsilon}^2)$.

CEMFI WORKING PAPERS

0601 *Beatriz Domínguez, Juan José Ganuza and Gerard Llobet*: "R&D in the pharmaceutical industry: a world of small innovations".

0602 *Guillermo Caruana and Liran Einav*: "Production targets".

0603 *Jose Ceron and Javier Suarez*: "Hot and cold housing markets: International evidence".

0604 *Gerard Llobet and Michael Manove*: "Network size and network capture".

0605 *Abel Elizalde*: "Credit risk models I: Default correlation in intensity models".

0606 *Abel Elizalde*: "Credit risk models II: Structural models".

0607 *Abel Elizalde*: "Credit risk models III: Reconciliation reduced – structural models".

0608 *Abel Elizalde*: "Credit risk models IV: Understanding and pricing CDOs".

0609 *Gema Zamarro*: "Accounting for heterogeneous returns in sequential schooling decisions".

0610 *Max Bruche*: "Estimating structural models of corporate bond prices".

0611 *Javier Díaz-Giménez and Josep Pijoan-Mas*: "Flat tax reforms in the U.S.: A boon for the income poor".

0612 *Max Bruche and Carlos González-Aguado*: "Recovery rates, default probabilities and the credit cycle".

0613 *Manuel Arellano and Jinyong Hahn*: "A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects".

0614 *Manuel Arellano and Stéphane Bonhomme*: "Robust priors in nonlinear panel data models".

0615 *Laura Crespo*: "Caring for parents and employment status of European mid-life women".

0701 *Damien Geradin, Anne Layne-Farrar and A. Jorge Padilla*: "Royalty stacking in high tech industries: separating myth from reality".

0702 *Anne Layne-Farrar, A. Jorge Padilla and Richard Schmalensee*: "Pricing patents for licensing in standard setting organizations: Making sense of *FRAND* commitments".

0703 *Damien Geradin, Anne Layne-Farrar and A. Jorge Padilla*: "The *ex ante* auction model for the control of market power in standard setting organizations".

0704 *Abel Elizalde*: "From Basel I to Basel II: An analysis of the three pillars".

0705 *Claudio Michelacci and Josep Pijoan-Mas*: "The effects of labor market conditions on working time: the US-UE experience".

0706 *Robert J. Aumann and Roberto Serrano*: "An economic index of *riskiness*".

0707 *Roberto Serrano*: "El uso de sistemas dinámicos estocásticos en la Teoría de Juegos y la Economía".

0708 *Antonio Cabrales and Roberto Serrano*: "Implementation in adaptive better-response dynamics".

0709 *Roberto Serrano*: "Cooperative games: Core and Shapley value".

0710 *Allan M. Feldman and Roberto Serrano*: "Arrow's impossibility theorem: Preference diversity in a single-profile world".

0711 *Victor Aguirregabiria and Pedro Mira*: “Dynamic discrete choice structural models: A Survey”.

0712 *Rene Saran and Roberto Serrano*: “The evolution of bidding behaviour in private-values auctions and double auctions”.

0713 *Gabriele Fiorentini and Enrique Sentana:* “On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models”.

0714 *Antonio Díaz de los Ríos and Enrique Sentana:* “Testing uncovered interest parity: A continuous-time approach”.

0715 *Francisco Peñaranda and Enrique Sentana:* “Duality in mean-variance frontiers with conditioning information”.

0716 *Stefano Gagliarducci, Tommaso Nannicini and Paolo Naticchioni:* “Electoral rules and politicians' behavior: A micro test”.

0717 *Laura Hospido*: “Modelling heterogeneity and dynamics in the volatility of individual wages”.

0718 *Samuel Bentolila, Juan J. Dolado and Juan F. Jimeno*: “Does immigration affect the Phillips curve? Some evidence for Spain”.

0719 *Enrique Moral-Benito*: “Determinants of economic growth: A Bayesian panel data approach”.

0801 *David Martinez-Miera and Rafael Repullo*: “Does competition reduce the risk of bank failure?”.

0802 *Joan Llull*: “The impact of immigration on productivity”.

0803 *Cristina López-Mayán:* “Microeconometric analysis of residential water demand”.