THE DEMAND FOR BOOKS AND OTHER PERIODIC PUBLICATIONS IN SPAIN

MANUEL JAÉN-GARCÍA
Department of Applied Economics. University of Almeria, Spain
mjaen@ual.es

ABSTRACT

This paper introduces a model for the demand of books and other periodic publications (BPP) in Spain based on micro-data from household budget surveys corresponding to 2006, 2007, and 2008. The model is formulated with a truncated or censored dependent variable in which selection is made using a probit model that determines the possibility of whether a given item is purchased or not. Then, based on the results from the initial model, the factors that determine expenditure on BPP are identified. Alternatively, a tobit model has been used. The paper is mainly focused on identifying the price and income elasticity of demand, as well as the influence of various demographic factors such as sex and age, both on the probability of purchase as well as on the expenditure made on BPP.

KEYWORDS: BOOK DEMAND; EXPENDITURE ON BOOKS, INCOME ELASTICITY, PRICE ELASTICITY.

RESUMEN

En este trabajo se presenta un modelo de demanda de libros y otras publicaciones periódicas (LPP) en España basado en los microdatos de las encuestas de presupuestos familiares (EPF) correspondientes a los años 2006, 2007 y 2008. El modelo se formula como un modelo de variable dependiente truncada o censurada en el que se realiza la selección con un modelo probit que nos determina la posibilidad de comprar o no el bien correspondiente y a partir de los resultados del modelo anterior se identifican los determinantes del gasto en LPP. De forma alternativa se utiliza un modelo tobit. El principal interés del trabajo está en identificar las elasticidades precio y renta de la demanda así como la influencia de diversos factores demográficos tales como sexo y edad tanto en la probabilidad de compra como en el gasto realizado en los bienes considerados.

PALABRAS CLAVE: DEMANDA DE LIBROS, GASTO EN LIBROS, ELASTICIDAD RENTA, ELASTICIDAD PRECIO.

INTRODUCTION

In Spain, the Book Law (Law 10/2007) regulates book prices in such a way that the maximum possible discount in department stores, bookstores and other points of sale is 5%. This legislation clearly favours publishing companies and distributors, consequently provoking distortion in the market that could be measured with the loss in consumer surpluses brought on by this situation. As a result, the welfare costs of this fixed price policy dramatically depend on price elasticity (Ringstad and Loyland, 2006; Beck, 2003; Ringstad, 2004). Therefore, a price variable was introduced into the model by means of the book price index from the consumer price index (CPI). This makes it possible to measure the price elasticity, which is expected to have a negative value. Nevertheless, it appears that income elasticity is of greater importance, and it also offers greater possibilities of making precise measurements.

In an initial approximation it seems that books, especially those that are not textbooks, are considered luxury goods in Spanish society, which is why we can expect an elevated positive correlation between demand for books and income and, consequently, a positive elasticity close to or even greater than one. Moreover, the greater the income elasticity, the higher the probability the book market will have a greater share in the national income over time.

It is the intention of this paper to contribute to the scarce amount of published material dedicated to the study of demand for printed books (Watt, 2007). Along with others, it is possible to quote the research of Bittlingmayer (1992), who found price elasticity between -2 and -3 for the German book market. He takes into consideration the demand for specific titles in that market. The data are obtained from publishers of a wide range of specialized books, and the analysis is not necessarily representative of books in general. The analysis covers the years 1984-86, but it only includes a selection of books published during those years. The main explanatory variable is book price and its principal result is price elasticity between -2 and -3. Most of the differences in demand are caused by other factors, which reflects the uncertainty of book demand and implies a corresponding economic risk when printing books. Hjorth-Andersen (2000) propose an equation for the aggregate demand in the Danish book market with four explanatory variables: book price, aggregate disposable income, number of titles, and a tendency measured by the number of years during the period studied, 1973-91. The study also concludes that book demand is sensitive to price, with an elasticity estimated around -1,4. It makes sense that the aggregate demand is less elastic to price than the demand for specific titles proposed by Bittlingmayer (1992). Increasing the price of a particular title reduces its demand, but it increases the demand for other books, so the total demand for books is less affected. These studies find that the income elasticity is around 1.8, meaning books are luxury goods. The number of titles does not appear to affect demand. On the other hand, there is a significant negative tendency in book demand, perhaps due to the crowding-out effect of new and cheaper substitutes close to books.

Ringstad and Loyland (2006) highlight several important ideas related to price and income elasticity of the demand for books. As regards price elasticity, they believe it is important for three reasons. Firstly, it supposes that there is monopolistic competition in the book market as there are many titles competing and fixed costs are very high. This implies that production is adjusted to a level of demand where price elasticity demand is less than -1. Secondly, if demand is sensitive to price, this would imply that subsidies are a possibly efficient means of increasing book purchases. Finally, the welfare costs for fixed book prices used in several European countries, Spain included, also crucially depend on price elasticity. As for income elasticity, it also matters a great deal as the sensitivity of income to demand for books combined with a positive general trend in income among consumers is decisive in determining the size of the book market. The greater the income elasticity, the higher the probability that the book market will represent a larger proportion of the GDP over time. Furthermore, books compete with a series of goods that are more or less related to them. The effect of the changes in the price of these goods on the demand for books depends on the size of the cross-price elasticity and on how significant the price changes actually are. The relative price of probable substitutes such as television, videos, or videogames has drastically diminished over recent decades, which could result in substantial crowding-out effects on the demand for books.

Prieto-Rodriguez et al. (2004), in their study of tax cuts on cultural goods, analyze cross-price elasticity and expenditure elasticity of 19 different groups of consumer goods including cultural goods. Their research is based on four-monthly data from the continuous family budget survey for 3 200 families for the period 1985-95 and a consumer price index. For books and other periodic publications they find an income elasticity of around 1,37 and a price elasticity of -1,65. The cross elasticity is negative for other cultural goods, -0,524 for movie, theatre, and shows and -0,331 for records and films, which would indicate that they are complementary goods to book purchases.

Also related to the Spanish market is the work of Palma et al. (2009), which analyzes price and income elasticity based on an aggregate model of temporal series for the period 1989-2006. They use the model applied by Hjorth-Andersen (2000) expanded with other economic variables. The main variables utilized are disposable income per inhabitant, average book price, price of substitute or alternative goods and a trend variable. Although they set out various models, among which one is based exclusively on cultural variables, they make the estimate for three models based on the two economic variables, the number of published titles and advertising. They find that the demand for books is elastic to income (1,19) and inelastic to price (between -0,61 and -0,75).

Villaroya and Scardibul (2010) use a static model for Spain in the same line that this one. They consider the effects of several factors, such as family, socio-economic and geographical environment, on household consumption of books and periodicals as well as on the amount spent. They find that education and occupational status of main-wage earner, household income and place of residence are important determinants for book and periodicals demand.

Gaffeo et al. (2008) utilize a dynamic model for information transfer in the Italian book market. They find that for each one of the three broad submarkets in which the Italian publishing industry can be classified (Italian and foreign novels and essays), sales over the course of the three-year sample can be adjusted by a Pareto distribution. The results can be interpreted in terms of an interaction model between buyers exchanging information about the books they purchase. These results support the principles of 'nobody knows', 'the winner takes it all' and 'success follows success', characteristics that the publishing industry shares with other creative industries. Moreover, the degree of uncertainly in the book market follows a seasonal trajectory with Christmas as the riskiest time of the year.

This paper utilizes a static approach from the individual consumer's point of view while taking into consideration the micro-data from the Household Budget Survey (HBS) conducted in Spain. This approach, in line with the study by Ringstad and Loyland (2006), makes it possible to employ a selection equation in order to elucidate what the determining economic and socioeconomic factors of book purchase probability are, and which factors determine expenditure on books. The HBS from 2006, 2007, and 2008 have been utilized, which together provide data for 63,054 families. This study has also analyzed data on prices from the CPI corresponding to the various Spanish provinces and different years. In this way, consumers from the same province in the same year face the same price indexes[1].

Theoretical fundamentals

In order to explain expenditure on books and other periodic publications (as of now referred to as 'BPP') based on individual expenditure contained in the micro-data, a sample selection model is implemented in line with the work of Heckman (1979). This paper follows in the steps of previously-published work by focusing on en-

1. Data was taken from the National Statistics Institute website, and others were provided by this organization.

dogenous switching regression models to model the selection and the desired quantity of BPP (Charlier et al., 2000; Lee and Trost, 1978; Olsen, 1980; and Jaén and Molina, 1994). A two-step switching model has been adapted to the cultural market, in particular, to expenditure on BPP.

For family j the model is specified as follows:

$$I_j^* = Z_j\alpha + \varepsilon_j \qquad [1]$$

$$Q_{Lj} = X_{oj}\beta_0 + \varepsilon_{Lj} \qquad [2]$$

with $(\varepsilon_j, \varepsilon_{Lj})$ following a bivariate normal distribution. The first sample selection equation makes it possible to differentiate those individuals for whom the variable Q_{Lj} was not observed. The second regression equation makes it possible to explain the variable of interest only for those individuals that had made a positive expenditure.

In the case analysed, Q_{Lj} is the expenditure on BPP made by the family, and I_j^* is an unobservable indicator that determines whether the family consumes BPP or not. Considering there is sample separation between families that purchase BPP and those that do not, Q_{Lj} will only be observed in one place and at one moment in time. In this way, whether a family purchases BPP or not is determined by I_j^*. That is, Q_j-Q_{Lj} if, and only if, I_j^* >0. The decision to purchase BPP or not and how much to spend on it are not independent decisions. Consequently, except in the case that ε_{Lj} are independent of ε_j, the standard estimation procedure by means of ordinary least squares produces biased estimators for the parameters, thus $E(\varepsilon_{Lj} | Z_j\alpha > \varepsilon_j) \neq 0$.

If the sample separation is identified by a dichotomic variable I_j, that is, $I_j = 1$ if $Z_j\alpha > \varepsilon_j$; and $I_j = 0$ in other case, and it is supposed that ε_j is normally distributed, the relationship is, in fact, the following probit model:

$$I_j^* = Z_j\alpha \quad \varepsilon_j \qquad [3]$$

with $I_j = 1$ if, and only if, I_j^* >0; and $I_j = 0$ in other case.

It is supposed that ε_j and ε_{Lj} follow a normal bivariate distribution with a zero mean and a non-singular common covariance matrix for the observations.

$$\Sigma = \begin{bmatrix} \sigma_L^2 & \sigma_{Lj} \\ \sigma_{jL} & 1 \end{bmatrix} \qquad [4]$$

It was therefore supposed, as can be observed in the matrix, that $\sigma_\varepsilon^2 = 1$.

It is easy to demonstrate[2] that:

$$E(\varepsilon_{Lj} | I_j = 1) - E(\varepsilon_{Lj} | I_j^* > 0) - E(\varepsilon_{0j} | Z_j\alpha > \varepsilon_j) - \sigma_{L\varepsilon} f(Z_j\alpha)/F(Z_j\alpha) \quad [5]$$

where *f* and *F* are respectively the functions of density and distribution of *N(0,1)* and

$$E(\varepsilon_{Lj}^2 | I_j = 1) = \sigma_0^2 - \sigma_{L\varepsilon}^2[(f(Z_j\alpha)/F(Z_j\alpha)(Z_j\alpha + f(Z_j\alpha)/F(Z_j\alpha)] \quad [6]$$

This then results in the following:

$$E(Q_{Lj} | I_j = 1) = E(Q_{Lj} | Z_j\alpha > \varepsilon_j) = X_{Lj}\beta_0 - \sigma_{L\varepsilon} f(Z_j\alpha)/F(Z_j\alpha) \quad [7]$$

And, consequently, the demand equation is specified in the following way:

$$Q_{Lj} = X_{Lj}\beta_0 - \sigma_{L\delta} f(Z_j\alpha)/F(Z_j\alpha) + \eta_{Lj} \qquad [8]$$

with $E(\eta_{Lj} | I_j = 1) = 0$.

With these equations the estimation procedure can be conducted in two stages. In the first, it is necessary to estimate the probit model where $pr(I_j \neq 1) = F(Z_j\alpha) + U_j$, with U_j as an error term in order to obtain consistent parameter estimates of a. Given that the probit function is concave, the maximum likelihood estimators can be obtained easily.

In the second stage, a is substituted by $\hat{\alpha}$ in the previous non-linear equation:

$$Q_{Lj} = X_{Lj}\beta_0 - \sigma_{L\delta} f(Z_j\hat{\alpha})/F(Z_j\hat{\alpha}) + \eta_{Lj} \qquad [9]$$

2. See Appendix 1

With the sample observations for people that purchase BPP, is it possible to estimate $\beta_0 \gamma \sigma_{L\varepsilon}$.

An alternative approximation to the previous procedure is to utilize the tobit method. This implies the following specification:

$$Q^*_{Lj} = X_{Lj}\gamma_0 + \mu_{ij} \quad [10]$$

where μ_{ij} is the error term,

$$Q_{Lj} = 0 \text{ if } Q^*_{Lj} \leq 0 \quad [11]$$

$$Q_{Lj} = Q^*_{Lj} \text{ if } Q^*_{Lj} > 0 \quad [12]$$

As for the empirical estimation, we used the linear logarithmic form for the equation of probability of purchase and the equation of demand, and for the Tobit equation. The exogeneous variables included in both equations are family income, the price indexes of BPP, the price indexes of other cultural goods and the price indexes for the rest of all other goods, as well as various sociodemographic variables. In this way, the probit equation is formulated in the following manner:

$$I^*_j = F_0 + F_1 \ln(Y_j) + F_2 \ln(P_{BPP}) + F_3 \ln(P_{SUST}) + F_4 \ln(P_{REST}) + \Sigma F_{5+i} Z_{ij} + e_j \quad [13]$$

where Y_j is income; P_{BPP} is the consumption price index of the subgroup of BPP corresponding to the 52 provinces and two autonomous cities; P_{SUST} is the price index of substitute goods (other cultural goods such as recreation, sports and cultural services); P_{REST} is the price index of the rest of all other goods[3]; and Z_{ij} is a combination of demographic variables that are defined below. It is supposed that the disturbance term e_j has a normal distribution with a zero mean and a variance equal to one.

For the equation of demand the same linear-logarithmic form was adopted:

$$\ln Q_{Lj} = ß_0 + ß_1 \ln(Y_j) + ß_2 \ln(P_{BPP}) + ß_3 \ln(P_{SUST}) + ß_4 \ln(P_{REST}) + \Sigma ß_{5+i} Z_{ij} + e_{Lj} \quad [14]$$

where the endogeneous variable Q_{Lj} measures expenditure on BPP.

3. All prices are deflated by general price index.

The estimation is carried out by means of the two-stage procedure previously mentioned. In the first stage the probit equation is estimated. The parameters estimated are utilized to build estimators for the variable $\Lambda_{Lj} = f(\hat{I}^*_j)/F(\hat{I}^*_j)$, where *f(.)* and *F(.)* are, respectively, the functions of demand and distribution of the *N(0,1)*[4], and $\hat{I}^*_j$ is the estimated value in the selection equation. By including this variable in the equation of demand, it is possible to explain the impact of the correlation between the errors in the equations of demand and the error in the selection equation. In the second stage, consequently, the equation of demand is estimated by adding Λ_{Lj} to the previous equation. As previously seen, upon carrying out the estimate of these equations using ordinary least squares, the estimators obtained are unbiased and consistent. Furthermore, a test can be made of the coefficient of Λ_{Lj} in order to test whether there is a significant statistical correlation between the errors from the two stages.

Data and definition of variables

The main source of data in Spain for the purposes of this paper is the consecutive Household Budget Surveys (HBS) of the National Statistics Institute (NSI, or INE in Spanish). The primary objective of these surveys is to discover household consumption patterns. The term 'household' is understood as the person or group of people that together occupy a family dwelling, or part of it, and consume food stuffs and other goods using a common budget.

This paper examines the HBSs from 2006, 2007, and 2008[5]. The size of the sampling is approximately 24,000 households per year. Each house-

4. This expression is known as inverse Mills´ ratio.

5. This new HBS, began in January 2006, substitutes the Continuous Household Budget Survey (CHBS) system from 1997, which was conducted every three months from 1997 until 2005.

hold remains in the sampling for two consecutive years, renewing half of the sample each year. The survey provides annual information about various characteristics related to the living conditions of the households (geographic, household members, principal wage earner, main dwelling and another dwellings, income, etc.), as well as about the different annual consumption expenditures.

Consumption expenditures refer to both the monetary quantity that households allocate to the payment of specific goods and services of final consumption, and also to the value of goods obtained for personal consumption, by self-provision, in-kind salary, free or reimbursed meals, and rent imputed to the dwelling in which the household is found (when the household is the owner of the dwelling or it is ceded by other households or institutions). Expenditures are registered at the moment of acquisition, independently of whether the payment is made in full or on several different occasions. The classification utilized to codify the expenditures is COICOP/HBS[6], structured into twelve large groups.

As regards price variables, considering that the HBS does not include product prices in its data, the corresponding consumer price indexes (CPI) were utilized.

The classification of the CPI provided by the NSI includes, on a provincial scale, the groups corresponding to books, press and other periodical publications (BPP) and recreation, sports and cultural services (other cultural goods that we consider as substitute goods of BPP). In this way the price index corresponding to the province and year were applied for each family.

The variables utilized can be divided into economic variables and sociodemographic variables. Included among the first group are product prices, income, and expenditure made. Among the second are the variables corresponding to the head of family and variables related to household conditions.

Regarding price variables, the price indexes used were those that corresponded to the three types of expenditure considered deflated by the general index. Also, all prices were considered utilizing the year 2001 as a reference year.

The signs expected for the coefficients are negative for BPP and positive for other cultural goods and the rest of goods, although it is expected that expenditure on BPP is independent of expenditure on the rest of all other goods, meaning the coefficient should not be significantly different from zero.

In the case of family income, given that total expenditure is usually greater than total income, income is understood as the total expenditure. In order to use disposable income, cash or in-kind subsidies are added to the total expenditure and taxes paid are subtracted.

This study examines current income. The reasoning behind this is that, considering the BPP epigraph represents a small percentage of family income and is a purchase that is made and paid once, the expenditure depends on the income available at the moment. This datum is taken from the HBS.

As for these sociodemographic variables[7], the variables utilized are those available in the HBS that can influence the demand and consumption of books:

a) Age of head of family. Dichotomic variables are used examining different age groups.

b) Gender of head of family. Likewise a dichotomic variable indicating gender.

6. Classification of Individual Consumption by Purpose/Household Budget Survey. This is an international classification of consumption used by Eurostat.

7. Appendix 2 thoroughly details the variables utilized as well as their statistics.

c) Level of education of head of family, with dichotomic variables indicating university studies, secondary, or no studies.

d) Job of head of family. Dichotomic variables are used that depend on a reduced classification of different kinds of jobs.

e) Socioeconomic status, with dichotomic variables that express if the head of family is an employed, unemployed, or retired person.

Some of these variables probably correlate to income, meaning that if it is necessary, they will be eliminated from the estimated equations.

f) Number of children, with a dichotomic variable depending on whether the household has more or less than two children.

g) Number of students in the household.

h) Size of the town, with dichotomic variables depending on that size.

As the data come from the years 2006, 2007 and 2008, a variable for trend is included to differentiate the year of origin.

In the probit equation the dependent variable is the selection variable that will take a value of 1 if the family purchases BPP and a value of 0 if otherwise. In the demand equation, the dependent variable is expenditure on BPP.

Results of the empirical estimation

This section presents the results from the estimate with the probit model in which it is seen how different family characteristics, as well as BPP price and other variables influence the probability of purchasing BPP. Secondly, the results are presented for the functions of demand for books, both when it is considered as the second stage of the analysis and when the estimate is made directly via the Tobit model.

The first results are presented in Table 1.

Table 1

RESULTS OF THE ESTIMATED PROBIT MODEL

Variable	Coefficient	Std. Error	z-Statistic
C	61,4354	8,4959	7,2319
Ln(TOTEXP)	0,1999	0,0064	31,0476
Ln(P_{BPP})	-0,0968	0,2215	0,4370
Ln(P_{SUST})	0,1788	0,8175	2,1870
Ln(P_{REST})	-14,2575	1,8410	-7,7443
AGE1	-0,0883	0,0226	-3,4515
AGE2	0,3246	0,0226	14,3361
AGE3	0,3257	0,01879	
NSTUDENT	0,505586	0,0415	12,179
DEPCHILD	-0,46	0,036	-12,7992
STUDY1	0,387221	0,01957	19,784
STUDY2	0,25248	0,01386	18,2139
JOB1	0,1784	0,0225	7,8388
JOB2	0,2538	0,225	11,2434
JOB3	0,1612	0,020	8,0165
JOB4	0,0809	0,0168	4,7919
SEX	0,3014	0,0141	21,2985
STATUS1	0,1887	0,0219	8,5858
STATUS2	0,0822	0,0227	3,6209
TOWNSIZE1	0,1678	0,01321	12,7027
TOWNSIZE2	0,0212	0,0182	1,1673
TOWNSIZE3	0,1046	0,0166	6,275
TREND	0,0043	0,00756	0,5709
McFadden R^2	0,107313		
LR statistic	8 596,94		
Prob (LR statistic)	0,0000		
Number of observations	63,054		

Source: this study with data from HBS.

The results obtained reveal a high value for the LR statistic, which indicates the rejection of the null hypothesis that all of the coefficients, except for the constant, are equal to zero. McFadden's statistic shows a low value but not one that is inferior to that obtained in other studies.

In Table 2 there is a contingency chart of the predicted response against the value observed for the dependent variable. The two tables and the associated statistics represent the results classified based on the specified cut off that, in this case, is 0,67.

Table 2
EXPECTATION-PREDICTION EVALUATION FOR BINARY SPECIFICATION

Success cutoff: C = 0,67						
	Estimated Equation			**Constant Probability**		
	Dep=0	**Dep=1**	**Total**	**Dep=0**	**Dep=1**	**Total**
P(Dep=1)<=C	13 591	14 130	27 721	20 904	42 150	63 054
P(Dep=1)>C	7313	28 020	35 333	0	0	0
Total	20 904	42 150	63 054	20 904	42 150	63 054
Correct	13 591	28 020	41 611	20 904	0	20 904
% Correct	65,02	66,48	65,99	100,00	0,00	33,15
% Incorrect	34,98	33,52	34,01	0,00	100,00	66,85
Total Gain*	-34,98	66,48	32,84			
Percent Gain**	NA	66,48	49,13			
	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	8 798,66	12 133,00	20 931,67	6 930,21	13 973,79	20 904,00
E(# of Dep=1)	12 105,34	30 017,00	42 122,33	13 973,79	28 176,21	42 150,00
Total	20 904,00	42 150,00	63 054,00	20 904,00	42 150,00	63 054,00
Correct	8798,66	30 017,00	38 815,66	6 930,21	28 176,21	35 106,41
% Correct	42,09	71,21	61,56	33,15	66,85	55,68
% Incorrect	57,91	28,79	38,44	66,85	33,15	44,32
Total Gain*	8,94	4,37	5,88			
Percent Gain**	13,37	13,17	13,27			

Source: this study with data from HBS.

In the left-hand portion of the first section of the table, observations that have a predicted probability $p=1-F(x\beta)$ are classified according to whether they are above or below the specified cut-off value. In the right-hand portion, the observations are classified using p, the sample proportion of *y=1* observations. This probability, which is constant among individuals, is the computed value to estimate a model that includes only the constant term *C*. The correct classifications are obtained when the predicted probability is lower or equal to the cut-off and the observed value *y=0*, or when the predicted probability is higher or equal to the cut-off and the observed value *y=1*. In this case 13 591 of the observations for *Dep=0* and 28 020 for *Dep=1* are classified correctly by the model utilized. Overall, the model utilized correctly predicts 65,99% of the observations. The increase in the number of correct predictions obtained upon moving from the right side of the equation to the left side provides a measurement of the model's predictive ability. Measurements of increase are observed in absolute percentage increases, 32,84%, and as a percentage of the incorrect classifications in the constant probability model corrected by the equation, 49,13%. The model utilized improves in the predictions of *Dep=1* by 66,48% but performs rather poorly in the case of *Dep=0* (-34,98%). Overall, the equation utilized is 32,84% better in predicted responses than the constant probability model. This change represents an improvement of 49,13% over the correct prediction percentage, up from the 33,15% of the default model.

In the second section of the table, the right-hand side of the equation contains the computations of the expected number of observations *y=0* and *y=1* for an estimated model with only one constant and for the equation utilized. Among the 20,904 individuals with *y=0*, the number of observations expected with *y=0* in the model utilized is 8798,66. Among the 42 150 observations with *y=1*, the number of observations expected with *y=1* is 30 017. These numbers represent a total increase of 5,88 percentage points (13,27%) over the constant probability model.

Finally, a goodness-of-fit test is made. To do so, this study utilized Pearson χ^2 type tests from Hosmer-Lemeshow and Andrews. The underlying idea of both tests is to compare the expected values adjusted with the real values according to group. If these differences are large, the model is rejected as it provides an insufficient adjustment of the data. The first test groups the observations on the basis of predicted probability for *y=1*. The second is a more general test that groups the observations on a basis of any series or expression of series. The first statistic obtains a value of 1 002,5646 for a χ^2 (998) with a p-value equal to 0,4535, while the second obtains a statistic of 1 089,6743 for a χ^2 (1,000) with a p-value equal to 0,0248, revealing mixed evidence for the goodness of fit.

It can be observed in Table 2 that family income has a positive and highly significant impact on the probability of purchasing books. This is not the case for the price of books and substitute goods, which, although they reveal the signs expected (negative and positive, respectively), they do not appear to have any significant influence on purchase probability. What does have an influence is the price of all other goods, which moves in the same direction as the BPP price. In other words, a rise in the price of other goods decreases the probability of books purchase.

When examining sociodemographic variables, it is observed that when the head of family is older, the probability of BPP purchase increases. The same occurs with the level of education. Finally, population size is an important factor. The larger the population, the greater the probability of purchase. If the head of family is a man, the probability of BPP purchases also increases. A low number of dependent children has a negative influence on the probability of purchase, while an elevated number of students increases said probability. Expressed in other terms, the greater the number of dependent children or students in the family unit, the higher the probability of purchase.

Interpreting the coefficient values in this model is complicated due to the fact that estimated coefficients cannot be interpreted as the marginal effect of the independent variables. The marginal effect of the variable x_j on the conditioned probability is given as:

$$\delta E(y_i/x_i\beta)/\sigma x_j = f\,(x_i\beta)\beta_j$$

where *f(x)=dF(x)/dx* is the density function corresponding to *F* (in this case the density function of the distribution *N(0,1)*). It is observed that *β* is weighted by a factor that depends on the values of all the regressors in *x*. Therefore, the direction of the effect of a change in *x* depends only on the coefficient *β*. Positive values for *β* imply that increasing *x* will increase the probability of the response. Negative values imply the opposite.

A calculation is made for the marginal impact of economic variables utilizing the formulation above. If we examine the average values of all the variables, the marginal impact of the income variable is 0 064176; for the BPP price variable it is 0,031170; for the price of all other goods it would be -4,5885, while for the substitute goods variable it is 0,057548. These marginal effects are comparable to the slopes (Green, 2010).

Table 3 contains the results obtained from the demand equation of the model.

TABLE 3
RESULTS FOR DEMAND EQUATION IN THE PROBIT MODEL

Variable	Coefficient	Std. Error	t-Statistic
C	-30,233	11,7198	-2,5796
Ln(TOTEXP)	0,7586	0,0195	38,7065
$Ln(P_{BPP})$	-0,9752	0,2742	-3,5561
$Ln(P_{SUST})$	0,6251	0,0981	6,3694
$Ln(P_{REST})$	7,0000	2,5917	2,7008
AGE1	-0,2518	0,0324	-7,7658
AGE2	-0,1748	0,0460	-3,7968
AGE3	-0,2806	0,0435	-6,4465
NSTUDENT	0,09437	0,0440	2,1442
DEPCHILD	-00009	0,045	-0,0207
STUDY1	0,0323	0,0420	0,7687
STUDY2	-0,02173	0,0309	-0,7014
JOB1	-0,0133	0,33168	-0,4026
JOB2	0,0300	0,0355	0,8441
JOB3	0,0266	0,030624	0,8712
JOB4	-0,0245	0,0245	-1,0035
SEX	0,1141	0,03275	3,4843
STATUS1	-0,0631	0,0357	-1,7665
STATUS2	-0,0788	0,00345	-2,2833
TOWNSIZE1	0,0434	0,0221	1,9979
TOWNSIZE2	0,10156	0,0221	4,5840
TOWNSIZE3	-0,0051	0,223	-0,2297
TREND	-0,1148	0,0090	-12,8083
Λ	-1,2944	0,1872	-6,9120
R^2	0,3087		
F-Statistic	783,8955		
Prob (F-statistic)	0,0000		
Number of observations	42,150		

Source: this study with data from HBS.

The results obtained are along the line expected. The most important variables are all highly significant. Unlike Ringstad and Loyland (2006), the determination coefficient value is quite elevated explaining more than 30% of the variation of the independent variable. The statistic *F* indicates that the model is significant overall. The inverse Mills´ ratio coefficient is negative as the theory predicts, and significant as well, which indicates that if it had not been added in the OLS estimate, the estimators obtained would have been biased and inconsistent. As for the economic variables, the income elasticity is approximately equal to 0,76. The value obtained is low compared to that of Prieto-Rodriguez et al. (2004) and Palma et al. (2009), although these studies utilize aggregate data. Regarding the price elasticity, it is very close to the unit and closer to that of Prieto, although not as high[8]. This indicates that the demand has unit elasticity with respect to price, contradicting the common belief among authorities and booksellers. The elasticities for other goods have the signs expected indicating all other cultural goods are substitutes for BPP, unlike the results of Prieto-Rodriguez et al. (2004) but in line with Ringstad and Loyland (2006).

As regards the sociodemographic variables, it is observed that the number of students positively influences the demand for BPP, as does male gender of the head of family and the family's location in big towns. Neither the job nor the level of education of the head of family have an influence, yet their status does have a negative influence.

Table 4 displays the results obtained for the Tobit model.

8. The corresponding Wald test does not reject the null hypothesis of equality to 1 of the coefficient.

Table 4
RESULTS FOR THE TOBIT MODEL

Variable	Coefficient	Std. Error	z-Statistic
C	8,2437	10,3162	0,7991
Ln(TOTEXP)	0,8831	0,0077	114,2851
$Ln(P_{BPP})$	-0,8962	0,3740	-3,2701
$Ln(P_{SUST})$	0,6324	0,0981	6,4426
$Ln(P_{REST})$	-2,118	2,2315	-0,9492
AGE1	-0,2578	0,0324	-7,9514
AGE2	0,0766	0,0282	2,7171
AGE3	-0,0321	0,02454	-1,3077
NSTUDENT	0,1090	0,03273	3,3312
DEPCHILD	-0,2244	0,0327	-6,8628
STUDY1	0,2792	0,02216	12,6014
STUDY2	0,1556	0,0173	8,963
JOB1	0,1139	0,0275	4,1298
JOB2	0,1937	0,0265	7,3085
JOB3	0,1457	0,0253	5,7571
JOB4	0,0471	0,0222	2,1231
SEX	0,771	0,0175	4,4052
STATUS1	0,0814	0,0289	2,8110
STATUS2	0,0145	0,03178	0,4589
TOWNSIZE1	0,147597	0,01566	9,4223
TOWNSIZE2	0,1182	0,0220	5,3684
TOWNSIZE3	0,06470	0,0199	3,2465
TREND	-0,11221	0,0090	-12,4203
SCALE:C(25)	1,3163	0,0045	290,3227
Number of observations	42,150		

Source: this study with data from HBS.

In order to analyse the overall significance of the model variables it is possible to conduct a test of redundant variables or rather a Wald test of equality to zero of the coefficients. In the first case a value is obtained for the likelihood ratio statistic of 15 515,00 for a χ^2 (23) meaning the corresponding p-value is 0,000 and, as a result, the null hypothesis of equality to zero of the coefficients is rejected. For the Wald test, a statistical value is obtained that is equal to 18 753,61 for a χ^2 (22) meaning p-value is 0,0000 and, consequently, the null hypothesis of equality to zero of the model coefficients is rejected.

In order to obtain the marginal effects of the variables it is necessary to take into account that they are given by the following expression:

$$\delta \ln (Q_{Lj}) / \delta \ln (Y) = \beta\Phi(\beta' X_i/\sigma)$$

where Φ is the distribution function of the *N(0,1)* and σ is the standard deviation of the regression and equal to all other relevant variables.

In this case the values of β coincide with the elasticities, making it possible to directly compare the results with those obtained in the two-phase model. The values obtained are very close in both models excluding the value corresponding to all other goods. Subsequently, it can be stated that the income elasticity takes a value close to 0,8, while the price elasticity takes a value close to the unit and all other cultural goods are substitute goods for BPP with an elasticity value close to 0,65.

The results corresponding to the socioeconomic variables indicate greater consumption of BPP among middle-aged citizens than for all others. Similarly, a higher level of education of the principal wage earner indicates greater consumption. This also occurs for occupation. When consumers have occupations with higher levels of education, this implies greater consumption of books. Unlike Ringstad and Loyland (2006), this study finds that the male gender of the head of family, as well as a higher socioeconomic level, implies greater consumption. The factor of population is also important: the larger the town the family lives in, the greater their consumption of books. When there is a low number of students in the family, consumption of books is also lower. Similarly, the lower the number of dependent children, the lower the consumption of books.

Summary and conclusions

This study has analysed the variables that influence BPP purchase probability and demand. Initially, a two-stage model was utilized, in line with Heckman, which calculated the purchase probability and, conditioned to it, the demand for books. Then, a Tobit model was implemented. This took into consideration both economic variables, which make it possible to calculate values of the price and income elasticities, and sociodemographic variables.

The results obtained are in line with similar studies. The income elasticity is close to 0,8 and the price elasticity is close to -1. This suggests that the demand for books is sensitive to both price and income, although not for the highest amount obtained in other studies. As for the results related to sociodemographic variables, a positive influence is obtained for an older age of the head of family, a better job and socioeconomic status, as well as the number of children and students in the household. This also occurs with the size of the town: the larger the size, the higher the probability and consumption of BPP. The male gender for the head of family also positively influences purchase probability and consumption of books.

It must be emphasized that although the purpose of this study was to analyse the demand for books in Spain, the impossibility of obtaining disaggregate information related to book prices, restricted the analysis to utilizing the most aggregate BPP variable. It is possible that this aggregation may have biased to some extent the elasticity values. Nevertheless, it is believed that they are qualitatively valid.

REFERENCES

Beck, J. (2003). Monopoly versus oligopoly on the debate on resale price maintenance for books: a preliminary empirical result. Diplomarbeit.

Bittlingmayer, G. (1992). "The elasticity of demand for books, resale price maintenance and the Lerner Index". Journal of Institutional and Theoretical Economics, 48, pp.588-606.

Charlier, E., Melenberg, B. and Van Soest, A. (2000). "An analysis of housing expenditure using semiparametric cross-section models". Empirical Economics, 25, pp. 437-462.

Gaffeo, E., Scorcu, A. E. and Vici, L. (2008). Demand distribution dynamics in creative industries: the market for books in Italy. Discussion paper num. 4, Dipartimento di Economia, Universitá degli Studi di Trento, pp. 1-37.

Green, W. H. (2010). Econometric Analysis. 6th Edition. Prentince Hall.

Heckman, J. J. (1979). "Sample selection bias as a specification error". Econometrica, 47 (1), pp. 153-161.

Hjorth-Andersen, C. (2000). "A model of the Danish book market". Journal of Cultural Economics, 24, pp. 27-43.

Jaén, M. and Molina, A. (1994). "Un análisis empírico de la tenencia y demanda de vivienda en Andalucía". Investigaciones Económicas, XVIII(1), pp. 143-164.

Lee, L. F. and Trost, R. P. (1978). "Estimation of some limits dependent variable models with application to housing demand". Journal of Econometrics, 8, pp. 357-382.

Olsen, R. J. (1980). "A least squares correction for selectivity bias". Econometrica, 48, pp. 1815-1820.

Palma, M., Martin, J. and Jaén, M. (2009). "El mercado del libro en España 1989-2006. Un análisis económico". Estudios de Economía Aplicada, 27-1, pp. 225-251.

Prieto-Rodriguez, J., Romero-Jordán, D. and Sanz-Sanz, J. F. (2004). "Is a tax cut on cultural goods consumption actually desirable?. A microsimulation analysis applied to Spain". Paper Presented at the AEA Conference on the Econometrics for Cultural Goods, Padova, April, 22-23.

Ringstad, V. (2004). "On the cultural blessings of fixed book prices: facts or fiction?". International Journal of Cultural Policy, 10, pp. 351-365.

Ringstad, V. and Loyland, K. (2006). "The demand for books estimated by means of consumer survey data". Journal of Cultural Economics, 30, pp. 141-155.

Villaroya, A. and Scardibul, J. G. (2010) "La demanda de libros y publicaciones periódicas en España". Estudios de Economía Aplicada Vol 28-1 pps 1-22.

Watt, R. (2007). "An empirical analysis of the economics of copyright: how valid are the results of studies in developed countries for developing countries?". In World Intellectual Property Organization (WIPO) (ed.), The economics of intellectual property – Suggestions for further research in developing countries and countries with economies in transition, Geneva: WIPO.

Recibido: 07-09-2011

Aceptado: 20-01-2012

Appendix

Demonstration of equations [7], [8], [9] and [10]

In order to carry out the demonstration, it is necessary to remember the definition and properties of the truncated normal distributions:

It is herein stated that a random variable *X* has a double normal truncated distribution if its probability density function is:

$$1/\left(\sqrt{2\pi}\right)\sigma \exp\left[-1/2\sigma^2(x-\xi)^2\right]\left[1/\left(\sqrt{2\pi}\right)\sigma\int_A^B \exp\left[-1/2\sigma^2(t-\xi)^2\right]dt\right]^{-1}$$
$$=\sigma^{-1}f\left(x-\xi/\sigma\right)\left[F\left(B-\xi/\sigma\right)-F\left(A-\xi/\sigma\right)\right]-1$$

where *f* and *F* are, respectively, the density and distribution functions of a random variable *N(0,1)*, the high and low cut-off points are *A* and *B*. If *A* is replaced by -∞ or *B* by +∞, the distribution is "simply truncated" from above or from below, respectively.

The expected value of the random variable *X* is given by:

$$E(X)=\xi+\left[\frac{f(A-\xi/\sigma)-f(B-\xi/\sigma)}{F(B-\xi/\sigma)-(F(A-\xi/\sigma)}\right]\sigma$$

and the variance by:

$$\operatorname{var}(X)=\left[1+\frac{(A-\xi/\sigma)f(A-\xi/\sigma)-(B-\xi/\sigma)f(B-\xi/\sigma)}{F(B-\xi/\sigma)-F(A-\xi/\sigma)}-\left\{\frac{f(A-\xi/\sigma)-f(B-\xi/\sigma)}{F(B-\xi/\sigma)-F(A-\xi/\sigma)}\right\}^2\right]\sigma^2$$

In this study, ε_j is considered the random variable *N(0,1)* meaning that, upon truncating from above, its mathematical expectation will be given by:

$$E(\varepsilon_j \mid \varepsilon_j < \mathrm{Z}_j\alpha)=-f(\mathrm{Z}_j\alpha)/F(\mathrm{Z}_j\alpha)$$

while its variance will be:

$$\operatorname{var}(\varepsilon_j \mid \varepsilon_j < \mathrm{Z}_j\alpha)=\left[1+\frac{-\mathrm{Z}_j\alpha f(\mathrm{Z}_j\alpha)}{F(\mathrm{Z}_j\alpha)}-\left\{\frac{-\mathrm{Z}_j\alpha}{F(\mathrm{Z}_j\alpha)}\right\}^2\right]$$

Upon truncating from below, the following is obtained:

$$E(\varepsilon_j \mid \varepsilon_j > Z_j\alpha) = \frac{f(Z_j\alpha)}{1 - F(Z_j\alpha)}$$

$$\mathrm{var}(\varepsilon_j \mid \varepsilon_j > Z_j\alpha) = \left[1 + \frac{Z_j\alpha f(Z_j\alpha)}{1 - F(Z_j\alpha)} - \left\{\frac{f(Z_j\alpha)}{1 - F(Z_j\alpha)}\right\}^2\right]$$

On the other hand, if we take into consideration the random variables ε_{0j} and ε_j the latter truncated from above, with distributions of $N(0,\sigma_0^2)$ and *N(0,1)*, respectively, the combined density function is given by :

$$P_{\varepsilon_j\varepsilon_{0j}}(x_1, x_2) = 1/F(Z_j\alpha)\exp\left[-1/2(1-\rho^2)(x_1^2 - 2\rho x_1 x_2/\sigma_0 + x_2^2/\sigma_0^2)\right]$$

where $\varepsilon_j < Z_j\alpha$.

Using the fact that the conditional distribution of ε_{0j} given ε_j, is normal with an expected value

$$E(\varepsilon_{0j} \mid \varepsilon_j) = \rho\sigma_0\varepsilon_j$$

and variance

$$\mathrm{var}(\varepsilon_{0j} \mid \varepsilon_j) = \sigma_0^2(1-\rho)^2$$

the result is:

$$E(\varepsilon_{0j} \mid \varepsilon_j < Z_j\alpha) = \rho\sigma_0 \cdot E(\varepsilon_j \mid \varepsilon_j < Z_j\alpha)$$

$$E(\varepsilon_{0j}^2 \mid \varepsilon_j < Z_j\alpha) = E\left[E\left(\varepsilon_{0j}^2 \mid \varepsilon_j\right)\right] = E\left(\rho^2\sigma_0^2\varepsilon_j^2 + \left(1-\rho^2\right)\sigma_0^2\right) = \rho^2\sigma_0^2 E\left(\varepsilon_j^2\right) + \left(1-\rho^2\right)\sigma_0^2$$

Taking into account that

$$\rho = \sigma_{0\varepsilon}/\sigma_0\sigma_\varepsilon$$

the result is:

$$E(\varepsilon_{0j} \mid \varepsilon_j < Z_j\alpha) = \sigma_{0\varepsilon} - f(Z_j\alpha)/F(Z_j\alpha)$$

$$E(\varepsilon_{0j}^2 \mid \varepsilon_j < Z_j\alpha) = \sigma_0^2 - \sigma_{0\varepsilon}^2\left[f\left(Z_j\alpha\right)/F\left(Z_j\alpha\right)\left\{Z_j\alpha + f\left(Z_j\alpha\right)/F\left(Z_j\alpha\right)\right\}\right]$$

Analogous calculations are also made for the case, which considers the combined distribution of ε_{Rj} and ε_j with the latter truncated from below.

Appendix 2

Variables

Table A1 describes the variables used in the analysis.

Table A

DESCRIPTION OF VARIABLES

Variable name	Definition
Ij	If the household purchase books: =1; if it do not purchase: =0.
$\mathrm{Ln}(Q_{Lj})$	Log of expenditure on BPP of the household.
Ln(TOTEXP)	Log of total expenditure of the household.
$\mathrm{Ln}(P_{BPP})$	Log of the consumer price index for BPP deflated by general price index.
$\mathrm{Ln}(P_{SUST})$	Log of the consumer price index for substitute goods (recreation, sports and cultural services) deflated by general price index.
$\mathrm{Ln}(P_{REST})$	Log of the consumer price index for the rest of goods deflated by general price index.
AGE1	If head of family is 35 years or less: =1; in other case: =0.
AGE2	If head of family is between 35 and 50: =1; in other case: =0.
AGE3	If head of family is between 50 and 65: =1; in other case: =0.
NSTUDENT	If the number of students in the household is two or more: =1; in other case: =0.
DEPCHILD	If the number of dependent children in the household is less than two: =1; in other case: =0.
STUDY1	If head has completed university studies=1; in other case: =0.
STUDY2	If head has completed secondary education=1; in other case: =0.
JOB1	If head works as a manager in a company or in a public administration=1; in other case: =0.
JOB2	If head works as a technician or professional=1; in other case: =0.
JOB3	If head works as an administrative employee or in the service or trade sector =1; in other case: =0.
JOB4	If head works as a craftsman, a skilled worker, operator or assembler =1; in other case: =0.
SEX	If head is a man: =1; in other case: =0.
STATUS1	Socioeconomic status of the main person. If he/she is working: =1; in other case: =0.
STATUS2	Socioeconomic status of head. If he/she is retired: =1; in other case: =0.
TOWNSIZE1	If the household lives in a municipality with 100,000 inhabitants or more: =1; in other case: =0.
TOWNSIZE2	If the household lives in a municipality between 100,000 and 50,000 inhabitants: =1; in other case: =0.
TOWNSIZE3	If the household lives in a municipality between 50,000 and 20,000: =1; in other case: =0.
TREND	If it is 2006: =1; 2007: =2; or 2008: =3.

Source: this study.

Table A2 contains some descriptive statistics of the variables used in the analysis.

Table A.2
DESCRIPTIVE STATISTICS OF THE VARIABLES USED

Variable name	Mean	Standard-error	Median	Maximum	Minimum
Ij	0,668475				
$Ln(Q_{Ij})$	5,505	1,4301	5,6412	10,5247	-3,465542
Ln(TOTEXP)	10,1930	0,622	10,2307	12,9478	5,5142
$Ln(P_{BPP})$	4,5822	0,027211	4,58057	4,689511	4,5076
$Ln(P_{SUST})$	4,5785	0,0692	4,5704	4,8381	4,4066
$Ln(P_{REST})$	4,6089	0,003122	4,609063	4,620945	4,601965
AGE1	0,13195				
AGE2	0,336251				
AGE3	0,272952				
NSTUDENT	0,9687				
DEPCHILD	0,965997				
STUDY1	0,2417				
STUDY2	0,424748				
JOB1	0,15846				
JOB2	0,1862				
JOB3	0,1503				
JOB4	0,3869				
SEX	0,754353				
STATUS1	0,617169				
STATUS2	0,287579				
TOWNSIZE1	0,363672				
TOWNSIZE2	0,1180				
TOWNSIZE3	0,1504				
Number of observations	63,054				

Source: this study from HBS data.