E-COMMERCE USE IN SPAIN[1]

Leonel Cerno

Universidad Europea de Madrid

leonel.cerno@uem.es

Teodosio Pérez Amaral

Universidad Complutense de Madrid

teodosio@ccee.ucm.es

This paper analyzes from the demand side the factors that influence private e-commerce in Spain. We use econometric models for the data of a survey of 18,948 individuals for 2003, of which 5,273 are internet users.

First, we analyze the determinants of the decision to purchase or not to purchase in the Web taking into account the link between e-commerce and the access and use of the Internet service. Then, we characterize the e-consumer profile. The model suggests that the main factors that influence the decision to use e-commerce are the accessibility to the Net, income and gender.

Second we use specific models for the users of e-commerce to measure the determinants of the number of purchases and the expenditure in the Web. For the expenditure equation we use the algorithm RETINA, which improves the forecast ability. This model can be used to predict the expenditure of new users with different individual characteristics.

Keywords: Internet expenditure, e-commerce use, e-commerce demand, e-commerce expenditure, RETINA.

[1] The authors acknowledge the financial support of the Cicyt Project SEJ2004-06948 of the Spanish Ministry of Education.

1. INTRODUCTION

An important difference between virtual and conventional markets is that there are fewer frictions in the virtual markets. In turn, one way of creating market power by virtual traders is to reduce the search costs. Internet has increased substantially the readiness of information on prices and products, allowing the consumers to identify the best option and to improve its position with respect to the on-line suppliers. The e-consumers have strengthened their position as the e-commerce has developed.

The consumer’s behavior starting from the onset of e-commerce. is analyzed by Clemons, Hann and Hitt (1999), Combers and Patel (1997) and Brynjolfsson and Smith (1999)

The on-line markets seem more efficient in terms of prices and costs. However, some studies find substantial price dispersion (Bakos, 2001). This dispersion can be explained in part, by the heterogeneity of specific factors such as the trust or the knowledge of the website or the brand name.

The analysis of the friction level of the Internet markets can be made from two points of view: comparing the characteristics between the two types of markets or analyzing the behaviour within the electronic market. The present paper adopts the second approach.

In the electronic markets, the efficiency is measured in four dimensions: the price levels, the price dispersion, the elasticities and the costs of distribution and other inputs (Smith, Bailey and Brynjolfsson, 1999). Regarding the price creation, there are factors that have similar effect in price formation as in conventional markets. For example search costs put downwards pressure on the prices and intensify the differentiation by suppliers to try to keep prices above marginal costs (Peddibhotla, 2000).

In turn, the after-sale service for some types of goods or even the effect of the Web size in equilibrium could be treated with the traditional microeconomic tools. Using the traditional assumptions, but going beyond the common belief that the prices in Internet are low because the consumers can find them easily and cheaply. Shapiro and Varian (1999) analyze under what conditions this happens.

In the present paper we study the subset of e-commerce buying referred to individuals. We use an empirical approach for a sample of 5,273 e-commerce buyers in Spain in 2003. The social and demographic impacts of different factors are measured and its implications for different formulations of e-demand are estimated. In section two we analyze which factors influence the decision of buying or not buying in Internet. In section three we analyze the e-demand from two perspectives: number of purchases and expenditure. Section four contains some conclusions and suggestions for further research.

2. E-COMMERCE AND INTERNET USE

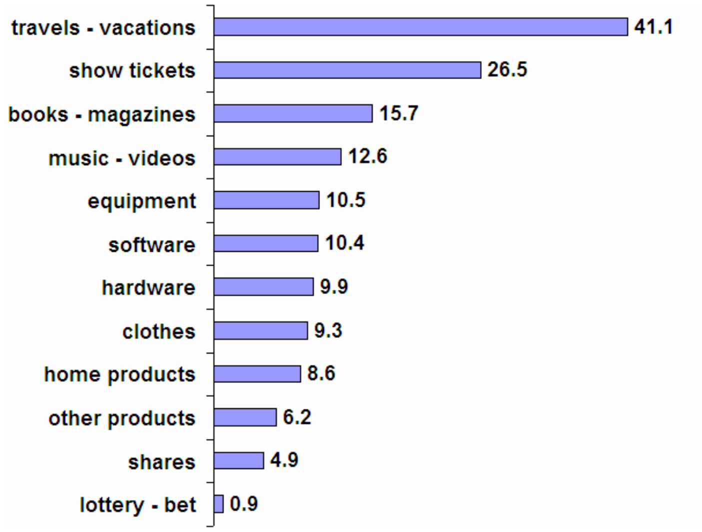
To understand the profile of the e-buyers we consider the determinants of the decision of buying or not buying through Internet. We use a binary choice model to measure the impact on the probability of using e-commerce of its determinants (Train, 2002). For explaining the behavior of the demand for e-commerce it is useful to consider the Internet access and use as explanatory variables (Cerno and Perez-Amaral, 2005).

2.1. DESCRIPTIVE ANALYSIS AND DEFINITION OF VARIABLES

The data are part of the survey TIC-H (2003) of the National Institute of Statistics (INE). It contains 5,273 individuals that are Internet users out of a total of 18,948.

In our dataset, only 3.7% buy through Internet in the last three months. In Graph 1 we have the percentages of the number of e-commerce goods and services in each of the categories:

Graph 1: Percentage of Individuals that purchase from each Category



We observe that in our sample the more demanded products are leisure activities such as travel (41.1%) and show tickets (26.5%) while stock trades and bets are only demanded by 4.9% and 0.9% of all e-buyers respectively.

2.1.1. CHARACTERISTICS OF THE DATA

Table 1. Definition of Variables

Variable	*Definition*
PURCH	Dummy =1 if the individual has bought through Internet
IS _ Q	Income quintile, sequential variable 1-5
HOMINT	Dummy =1 if the individual has Internet access at home
USE	Number of different access modes used (1 to 4)[2]
USAGEINT	Internet intensity of use (hours per week)
USAGECOM	Computer intensity of use (hours per week)
SEXM	Dummy =1 if the individual is male
AGE	Age of the individual
POPUL	Relative population size (provincial level)[3]
MEMB	Number of residents in the household

[2] The four places where the individual access to Internet are home, workplace, center of study and other places such as hotels, libraries, cybercafes, etc.
[3] Referred to the provincial population size of the individual divided by the total population in Spain.

Table 2. Summary Statistics

Variable	*Average*	*Standard Deviation*	*Minimum*	*Maximum*
PURCH	0.126	0.336	0	1
IS _ Q	4.030	0.850	1	5
HOMINT	0.615	0.487	0	1
USE	1.490	0.653	1	4
USAGEINT	38.001	29.805	0	70
USAGECOM	46.642	29.257	0	70
SEXM	0.519	0.499	0	1
AGE	33.862	12.465	15	88
POPUL	4.237	4.429	0.163	13.277
MEMB	3.333	1.192	1	6

Starting from this first analysis, we estimate a model for the e-buy determinants including economic and demographic attributes like the income, gender, age and habitat size. We also consider the individual's characteristics regarding the Internet service such as the access from the home, the use from other places besides the home and the intensity of use measured in hours per week. The binary logit model is:

$$\ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1\ IS\_Q_i + \beta_2\ HOMEINT_i + \beta_3\ USE_i + \beta_4\ USAGEINT_i +$$

$$+\beta_5\ USAGECOMP_i + \beta_6\ SEXM_i + \beta_8\ AGE_i + \beta_9\ AGE_i^2 + \beta_{10}\ POPUL_i + \beta_{11}\ MEMB_i + u_i$$

Where $P_i = \Pr\left(PURCH_i\right)$ is the probability that the respondent *i* buys through Internet in the last three months. Next we present the estimation results and the odd ratios:

Table 3. Logit results for the use or not use of e-commerce equation

Dependent Variable: $PURCH$	Coefficient	Odd Ratio	z
Constant	-6.19	---	13.67
IS_Q	0.23	1.26	3.75
$HOMEINT$	0.48	1.62	4.39
USE	0.39	1.48	5.84
$USAGEINT$	0.01	1.01	6.31
$USAGECOMP$	0.01	1.01	2.61
$SEXM$	0.38	1.46	4.24
AGE	0.09	1.09	4.01
AGE^2	-0.001	0.99	4.22
$POPUL$	0.04	1.04	4.77
$MEMB$	-0.08	0.92	2.16

Sample: 5,223

Log-likelihood: -1773.45

$\chi^2_{(10)}$: 397.09 (Prob. = 0.000)

Pseudo R^2: 0.1007

The global significance of the model is high since the log-likelihood 397.09, is highly significant for a chi-square with 10 degrees of freedom. The goodness of fit is also good according to the within sample predictions of Table 4:

Table 4. Goodness of fit. In sample predictions.

		Actual Value		
		$PURCH = 0$	$PURCH = 1$	*Total*
Predicted Value	$PURCH = 0$	3,063	208	3,271
	$PURCH = 1$	1,521	451	1,972
	Total	4,584	659	5,243

The model predicts correctly 3,514 of the 5,243 observations (67.02%). Due to the large number of zeros in the sample we use as threshold the proportion of zeros in the endogenous variable (0.126). So the specificity (percentage of correct “zeros” predicted inside the sample) is 66.81%, and the sensibility (percentage of correct “ones” predicted inside the sample) is 68.43%.

Since the sample is composed of individual private Internet users, the results only apply to this group. We conclude that:

- All the explanatory variables have positive effects in the probability of e-buying, except family size.
- The Internet access at home (*HOMEINT*) and the different access modes used (*USE*) refer to the individual Internet habits, and have high odd ratios. Internet access at home has the highest odd ratio of the two (1.62), followed by the different access modes used with 1.48. That is to say the Internet access at home contributes a 62% to the e-buy probability and the number of different access modes used contributes 48%.
- The male gender influences positively with an odd ratio of 1.46.
- The influence of income is positive with an odd ratio of 1.26.
- Population habitat (*POPUL*), Internet intensity of use (*USAGEINT*) and computer intensity of use (*USAGECOMP*) will also have positive impacts in the e-buy probability, but less so than those above. The odd ratios are close to 1 (1.04 for the *POPUL*, and 1.01 both for the variables *USAGEINT* and *USAGECOMP*).
- The case of *AGE* is special since it is included in the model both in levels and in squares. This is a common specification in applied econometrics. It is a quadratic function in *AGE*. Its first derivative is a linear function in *AGE*: $\partial PU\hat{R}CH / \partial AGE = 0,09 - (2).(0,001)AGE$. That allows that at early ages, the probability increases with age, but at a diminishing rate. For older ages, when age increases the probability decreases. The maximum effect is at 45 years.
- The only variable that diminishes the probability is the household size, *MEMB*. Its odd ratio is 0.92, the only one below 1

Last we evaluate the predicted probability variation by levels of income, age, connection at home and number of places of Internet use.

Table 5. Probability of Internet Purchase by Levels of Income

Income Quintile	1º	2º	3º	4º	5º
Probability	0.0815	0.1004	0.1232	0.1503	0.1821
Difference		0.0189	0.0228	0.0271	0.0318

Table 6. Probability of Internet Purchase by Places of Internet Use

Places of Internet Use	1	2	3	4
Probability	0.1283	0.1786	0.2431	0.3217
Difference		0.0503	0.0645	0.0786

Table 7. Probability of Internet Purchase by Internet Access at Home

Internet at Home	No	Yes
Probability	0.1171	0.1765
Difference		0.0479

Of these results we can say:

- The higher increment in the probability is when the individuals use the Internet in several places, being of 7.86% when going from three to four places.
- Changing income quintile doesn't always provoke the same increment in the e-buy probability. The probability differences increase with income.
- Having Internet at home increases the probability of purchase in the Web, with a higher increment than the differences estimated for the different income categories.

Therefore, as a conclusion we can say that, of all the determinants for purchasing through the Internet, the more important will be the characteristics of the individual like Internet access at home and the places of use besides the home. Surprisingly the Internet use intensity (weekly hours connected to the Web) has a positive impact but smaller than Internet access at home and places of Internet use. The other two variables that have positive impacts in the probability are the male gender and income level, both with important but smaller impacts.

Once we have analyzed the decision of to buy or not to buy in internet, we study the determinants of the next decisions: first how many times to purchase and second how much to spend.

3. SPECIFICATION OF THE DEMAND MODELS

Modeling demand or expenditure in telecommunications is not easy. Frequently the available data are binary and/or incomplete. Sometimes we do not have information on the prices paid by the consumers, or lack information on income or other variables. The functional form is usually unknown. Heterogeneity of the sample is frequent.

Before estimating e-commerce demand and expenditure functions, we describe the variables that will be used. Then the heterogeneity will be treated using a methodology called k-means (Mac Queen, 1967; Hartigan and Wong, 1979) to identify groups with similar characteristics in the sample.

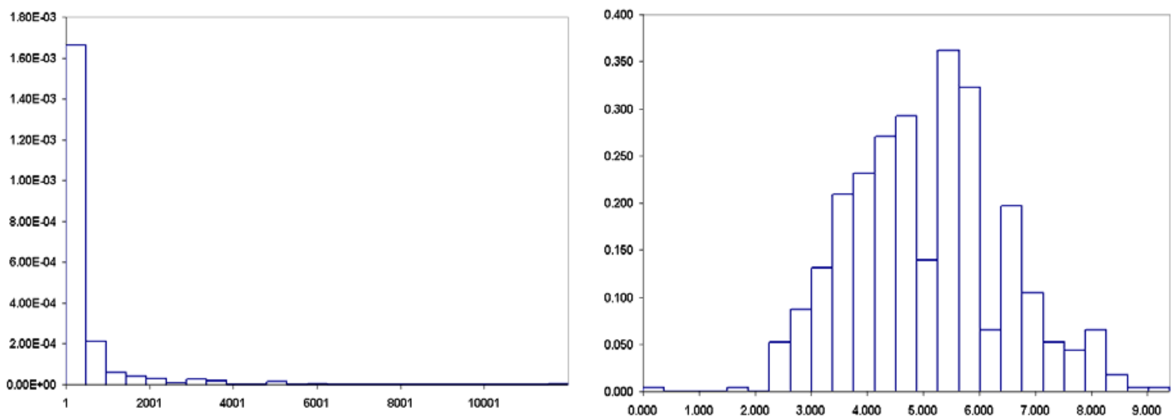
The additional variables in this section are in Table 8. Most of them are dummies. Notice that here the proxy for individual income is continuous, as originally constructed.

Table 8. Definition of Additional Variables

Variable	*Definition*
$QUANT_i$	Quantity of transactions for buy in Internet.
G_i[4]	Expenditure in Internet in the last three months (in euros).
IS_i	Individual Income Index[5].
$STUDYLEVEL_i$	years of schooling
P_1 to P_{12}	Dummies =1 if individual buys respectively home products, music, books, clothes, etc (see Graph 1)
$FP1_i$ to $FP4_i$	Dummies =1 if the payment is through credit card, bank transfer, against refund or subscription.

Next we make descriptive and graphic analyses of the variables comparing with the expenditures in the Web. Observing the histogram of the variable G_i both in levels and logarithms, the evidence of the outliers is clear:

Graph 3. Histogram of expenditures in internet G, in levels and logs



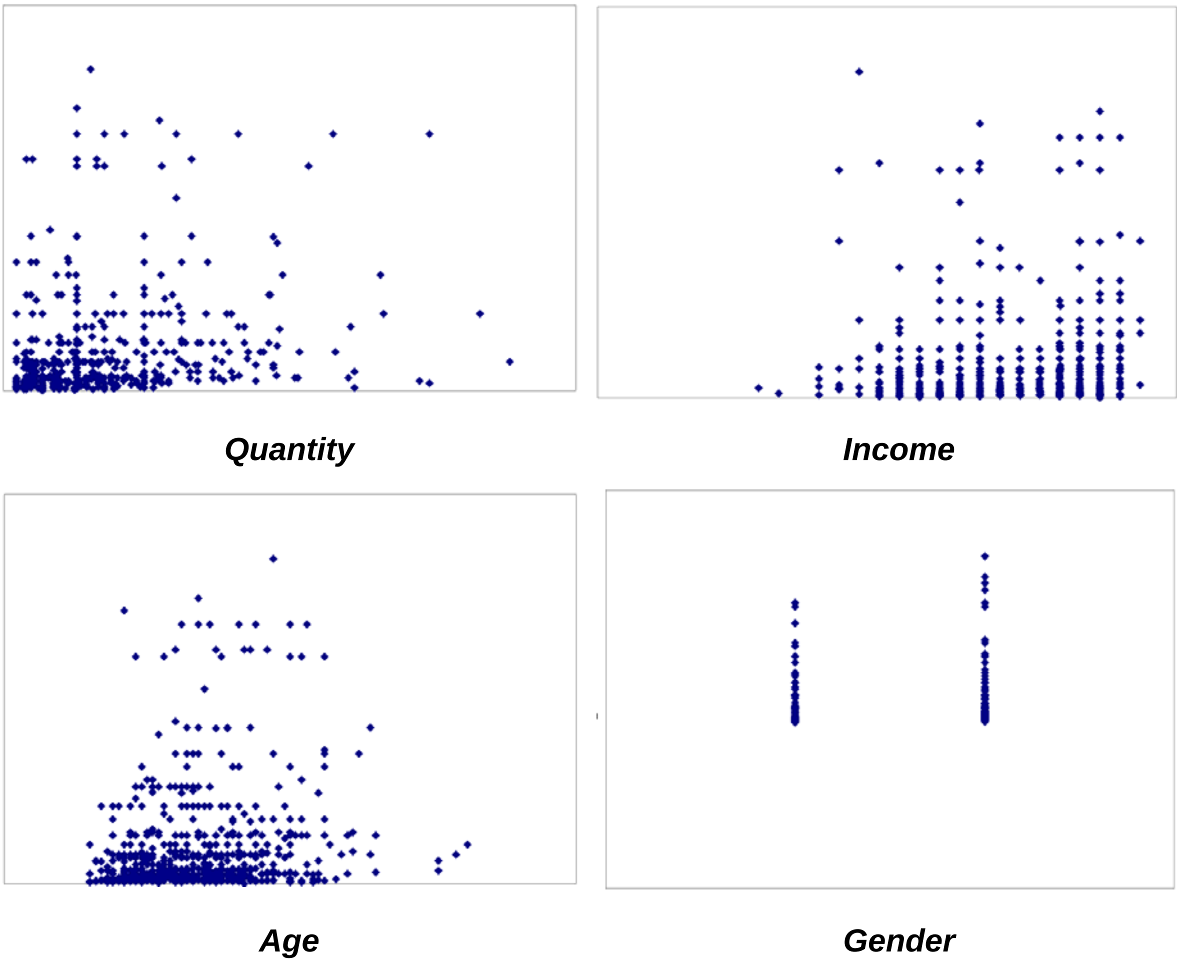
Levels ***Logs***

We use the algorithm of Peña and Yohai (1999) for detection and exclusion of extreme values. When applied to the logarithm of expense, it excludes 22 extreme values. Next we plot bivariate graphs with the expenditures in e-commerce versus selected candidate explanatory variables:

4 G_i are euros spent in products through Internet in the last three months.

5. Contructed at the individual level as weighted average of non human and human capital. See Cerno and Pérez Amaral (2005)

Graph 4. Expenditure vs. selected variables. The vertical axis is always Expenditure, *G*



The information was collected by INE using a combination of CATI[6] and PAPI[7] procedures. The data was subsequently filtered, so it is hard to imagine that any outlier is due to incorrectly registered data. However heterogeneity in the expenditure behaviour is likely to be present.

Next we characterize groups of individuals that have homogeneous profiles. For that we use the probability of an individual buying in internet, estimated in the previous section and the k-means algorithm of Mac Queen (1967), a classic method to detect homogeneous subgroups. The homogeneity criterion is to minimize the sum of squares within each group for all the variables.

[6] CATI: Acronym of Computer Assisted Telephonic Interview.
[7] PAPI: Paper and Pencil interview.

$$\min SSG = \sum_{g=1}^{G}\sum_{j=1}^{P}\sum_{i=1}^{n_g}\left(x_{ijg} - \overline{x}_{jg}\right)^2 = \min \sum_{g=1}^{G}\sum_{j=1}^{P} n_g s_{jg}^2$$

Where *SSG* is the Sum of Squares of Groups, x_{ijg} is the value for individual i of variable j classified in group g, and $\overline{x}_{jg}$ is the mean of this variable within the group. It can be shown that this is equivalent to minimizing the weighted sum of the variances of the variables in each group. These group variances can be considered a sensible measure of heterogeneity. The k-means algorithm follows these steps (Peña, 2002):

a) Select the k points that are further between themselves as centers of the initial groups.
b) Compute the Euclidean distances between each element and the centers of each of the k groups. We assign each element to the group whose center is closest.
c) Define the homogeneity criterion (in our case to minimize the *SSG* and continue reassigning.
d) If after the reasignation the criterion does not improve, then end the process.

We have classified the sample in five groups, based on the probability intervals calculated from model (4.4). This could change depending upon the level of aggregation and homogeneity. We have the following groups:

Table 9. Internet Users by Probability of purchase in the Internet.

GROUP	Probability of purchase in the Internet	Group size
Less propense (Group 1)	0.010 – 1.101	1,757
Little propense (Group 2)	0.101 – 0.191	1,341
Average propensity (Group 3)	0.191 – 0.298	1,037
High propensity (Group 4)	0.299 – 0.429	684
Very high propensity (Group 5)	0.430 – 0.701	411

We treat these groups by using group specific constants in the models of demand for e-commerce.

3.1. MODEL OF DEMAND THROUGH THE INTERNET.

Here we estimate the parameters of a demand for the number of purchases of goods and services in Internet. The endogenous variable $QUANT_i$, is the number of purchases performed via internet by an individual. Since the data are natural numbers we use a Poisson model. Here $QUANT_i$ is a function of income, the four forms of payment, gender, age, education and population size. We treat the heterogeneity using specific dummies for each of the groups of Table 9. Now the model is

$$QUANT_i = \alpha_0 + \alpha_1 IS_i + \alpha_2 STUDYLEVEL_i + \alpha_3 SEXM_i + \alpha_4 AGE_i +$$

$$+\alpha_5 POPUL_i + \sum_{k=1}^{4} \phi_{ik} FP_k + \sum_{m=1}^{5} \xi_{im} GR_m + u_i$$

where $i = 1,\ ...,\ 5,218$. These individuals are the ones that purchase in Internet. The variables are as described above. FP_k refers to the four modalities of payment considered, while GR_m refers to the groups by probability of purchase.

3.2. RESULTS AND DISCUSSION

Table 10. Frequency of use of e-commerce Poisson model

Dependent Variable	Coefficient	incidence ratio	z
QUANT			
Constant	-3.63	---	17.08
IS	0.42	1.51	4.67
STUDYLEVEL	-2.2	0.11	4.74
SEXM	0.12	1.13	1.93
AGE	0.002	1	0.77
POPUL	0.1	1.01	1.69
$FP1$	2.42	11.23	36
$FP2$	0.82	2.27	10.93
$FP3$	1.32	3.75	19.79
$FP4$	0.36	1.42	2.72
$GR1$	-0.79	0.45	7.69
$GR2$	-0.22	0.81	2.84
$GR3$	-1.14	0.32	5.14
$GR4$	0.11	1.01	0.13

Sample size : 5,218

Log-likelihood: -1,914.34

$\chi^2_{(13)}$: 3,635.19 (Prob. = 0.000)

Pseudo R^2: 0.4870

From Table 10 we have that:

- The expected number of purchases in Internet is mostly related to access to a credit card, ($FP1$).
- The incidence of the rest of means of payment is smaller. Payments via bank transfer ($FP2$) and postage due ($FP3$) influence by 2.27 and 3.75 respectively.
- Payment by subscription is also important, but its incidence is the smallest.

- The individual income *(IS)* has a positive influence in *QUANT.* Its incidence is 1.51. A unit increase in *IS* is associated with a 51% increase in *QUANT.*
- Other factors such as age, education, and population size seem to have little influence on the expected quantity of transactions in the Internet.

3.3. **PREDICTION OF THE EXPENDITURE IN E-COMMERCE.**

The forecast of expenditures is important for several reasons. It allows to evaluate the costs and benefits of different policy measures, and to compare different scenarios of evolution of consumer habits. In this section we seek to formulate a forecast model for the expenditures in e-commerce. For that we use a general linear model in which we treat the different groups with constant specific dummies. The dependent variable is the log of the expenditure in e-commerce.

$$\log\left(G_i\right) = \gamma_0 + \gamma_1 \log\left(IS_i\right) + \gamma_3 \log\left(AGE_i\right) + \gamma_4 SEXM_i + \sum_{g=1}^{4} \delta_{ig} GR_i + \sum_{p=1}^{12} \xi_{ip} P_i + u_i$$

The specific constants GR_i and P_i correspond to the five groups of individuals of the previous section and to the twelve types of products respectively.

Next we use the algorithm RETINA of Pérez Amaral, Gallo and White (2003) that chooses the model with the best out of sample predictive ability, in terms of its AIC (Akaike Information Criterion).

The models suggested by RETINA are:

1. **linear in the parameters** for quick and precise estimation,
2. **nonlinear in the inputs** (explanatory variables), to enhance its approximation capabilities. The regressors used by RETINA are squares, cross products and ratios of the original inputs,
3. **parsimonious** in the use of parameters for improved out of sample prediction capabilities and

4. can be thought of as a **generalization of ARIMA for cross section** data and other types of data.

Table 11 shows the transformations used by RETINA. More results in the Appendix

Table 11. Variables Used by RETINA

Endogenous Variable	$\ln(G)$
Original Continuous Variables	$\ln(IS)$, ln (*AGE*)
Specific Constants	*SEXM*, G_g where $g=1, \ldots, 5$
Specific Slopes	*SEXM*, G_g , P_p where $p=1, \ldots, 12$

GR_g is referred to the five groups of individuals previously detected and P_p to the twelve groups of goods and services that are traded in the Internet.

Table 12 shows the main statistics for comparing the basic linear model that use no interactions and the model chosen by RETINA. We notice an important decrease of the robust mean square prediction error (RCMSPE) from 1.195 to 0.689, which is indicative of better out of sample prediction ability.

Table 12: The Basic linear model and the RETINA models for the expenditure function

	GLM (BLM)	RETINA
Parameters	8	33
AIC	0,356	-0,813
$\bar{R}^2$	0,26	0,779
RCMSPE	1,195	0,689

GLM means General Linear Model, and BLM is Benchmark Linear Model. The within sample goodness of fit $\bar{R}^2$ increases from 0.260 to a high value of 0.779. From the results in the Appendix we also conclude that

- In the RETINA model the original inputs enter basically as transformations, that is interactions, squares and ratios. So the transformations seem to improve the prediction capabilities.
- Of all the groups of consumers the only one that shows up as significantly different in the RETINA model is the second one.
- Summarizing, RETINA suggests a model for predicting expenditure in e-commerce that is substantially better than the baseline linear model. Moreover it is easy to estimate and to use for prediction.

4. CONCLUSIONS

In this paper we analyze the demand for electronic commerce in Spain by individuals, using models that help us respond three questions.

1. What variables influence the decision to **use or not to use** e-commerce? Our first model analyzes the factors that influence the decision of each individual to buy or not to buy via internet.

The results of the first model tend to confirm that the main determinants of the buyer are his/her income level, age and study level. However, gender and population size also have influence.

The typical individual demanding goods and services through Internet in Spain in 2003 is a young person with higher education, living in a medium sized population, with median level of income that allows him access to credit and allows the use of credit cards.

2. What factors determine the **number of times that an individual uses** e-commerce? Our second model analyzes how much the characteristics of each individual influence the quantity of transactions performed through Internet.

For that we use a Poisson model and conclude that the access to the different forms of payment through the net (credit card, bank transfer, postage due and subscription) will influence the number of transactions. Individual income and several socioeconomic attributes are also relevant.

3. What are the determinants of **how much money is spent** in e-commerce by each consumer? Our third model is an expenditure function that relates the expenditure of each individual e-buyer to his/her economic and socio demographic characteristics.

The bivariate graphics of **expenditure** in e-purchases against its likely determinants reveal some anomalies and heterogeneity. We use the Yohai and Peña method for the outliers and the k-means algorithm to detect 5 groups of consumers.

Next we use a basic linear model together with models suggested by the new methodology of RETINA. We obtain new models that are linear in the parameters but non linear in the inputs with enhanced out of sample prediction capabilities and improved within sample fit. The coefficient of determination (in-sample goodness of fit) increases almost threefold, and the out of sample forecast ability increases substantially since the RCMSPE decreases by almost a half.

The types of models used here for cross section data with heterogeneity can be used in other studies for different years or regions in Spain or elsewhere.

One of the main goals of this line of research is to become a tool for the decision makers to enhance the development of the Information Society in Spain. For that reason future research will focus on analyzing more recent data such as the 2006 survey. This will allow comparisons over time, market by market, and the detection of regularities and changes in the behavior. Specific policy recommendations can be obtained when using our results in conjunction with specific scenarios and policy goals.

5. REFERENCES

Angrist, J. (2004) "Treatment Effect Heterogeneity in Theory and Practice" Economic Journal, Royal Economic Society, vol. 114 (494), pp. C52-C83.

Bakos, Y. (1997) "Reducing Buyer Search Cost: Implications for Electronic Marketplaces" Management Science, vol.43, Issue 12.

Bakos, Y. (2001) "The Emerging Landscape for Retail E-Commerce" Journal of Economic Perspectives, vol. 15, nº 1, pp. 69-80.

Brynjolfsson, E. and Smith, M (1999) "Frictionless Commerce? A Comparison of Internet and Conventional Retailers" Working Paper, MIT Sloan School of Management.

Cerno, L. and Pérez-Amaral, T. (2005) "Demand of Internet Access and Use in Spain" Documento de Trabajo 0506. Instituto Complutense de Análisis Económico (ICAE).

Cerno, L. and Pérez-Amaral, T. (2006) "Medición y Determinantes de la Brecha Tecnológica en España", mimeo, Universidad Complutense de Madrid.

Fundación AUNA (2003) "Informe España 2003", disponible en http://www.fundacionauna.com/areas/25_publicaciones/

Greene, W. (2000) "Econometric Analysis" 4th Edition, Prentice Hall

Hartigan, J. and Wong, M. (1979) "A k-means Clustering Algorithm" Applied Statistics, 28, 100-108.

Heckman, J. (1978) "Dummy Endogenous Variables in a Simultaneous Equation System" Econometrica 46: 931-959.

Hotelling, H. (1929) "Stability in Competition" The Economic Journal, pp. 41-57.

INE, "Encuesta Sobre Equipamiento y Uso de Tecnologías de Información y Comunicación de los Hogares del Segundo Trimestre de 2005" in www.ine.es/prensa/np388.pdf

Mac Queen, J. (1967) "Some Methods for Classification and Analysis of Multivariate Observations" Proc. Symposium of Mathematics, Statistics and Probability, 5th, pp. 281-297, Berkeley

Maddala, G. (1983) "Limited Dependent and Qualitative Variables in Econometrics" Cambridge University Press

Marinucci, M. and Pérez-Amaral, T. (2005) "Econometric Modelling of Business Telecommunications Demand Using RETINA and Finite Mixtures" Documento de Trabajo 0501. Instituto Complutense de Análisis Económico (ICAE).

Mas-Colell, A, Whinston, M. and Green, J. (1995) "Microeconomic Theory" Oxford University Press.

McKinsey, J. (2000) "Internet Pricing. A Creator of Value Not a Destroyer. McKinsey Marketing Practice" Documento disponible en http://marketing.mckinsey.com/solutions/McK-Internet_Pricing.pdf

Peddibhotla, N. (2000) "Are Economic Theories Good Enough to Understand E-Commerce?" en Wiseman, A. "The Internet Economy: Access, Taxes, and Market Structure", Brookings Institution Press, Washington D.C.

Peña, D. (2002) "Análisis de Datos Multivariantes", Mc Graw Hill.

Peña, D. and Yohai, V. (1999) "A Fast Procedure for Outlier Diagnostic in Large Regression Problems" Journal of the American Economic Association, 94: 434-445.

Pérez-Amaral, T. Gallo, G. and White, H. (2003) "A Flexible Tool for Model Building: the Relevant Transformation of the Inputs Network Approach (RETINA)" Oxford Bulletin of Economics and Statistics, 65:821-838.

Red.es (2004) – Ministerio de Industria, Turismo y Comercio "Las TIC en los Hogares Españoles" disponible en http://observatorio.red.es/estudios/consumo/index.html

Shapiro, C. and Varian, H. (1998) "Information Rules: A Strategic Guide to the Network Economy", Harvard Business School Press.

Smith, M., Bailey, J. and Brynjolfsson, E., (1999) "Understanding the Digital Markets: Review and Assessment" en Brynjolfsson, E. and Kahin, B. "Understanding the Digital Economy", MIT Press.

Train, K. (2002) "Discrete Choice Methods with Simulation" Cambridge University Press.

Varian, H. (1992) "Análisis Microeconómico" 3rd. Edition, Antoni Bosch ed.

Wilde, J. (2000) "Identification of Multiple Equation Probit Models with Endogenous Dummy Regressors". Journal of Econometrics 30: 415-443.

Zellner, A. (1962) "An Efficient Method of Estimating Seemingly Unrelated Regressions and Test of Aggregation Bias" Journal of the American Statistical Association, 57, pp. 500-509.

APPENDIX. Linear expenditure function

Sample size	614	
Adjusted R^2	0.26	
Standard error of estimation	1.199	
RCMSPE	1.195	
AIC	0.356	
Variable	**Coefficient**	***t* statistic**
Constant	2.38	3.96
Continuous original variables		
$\log(IS)$	0.17	0.65
$\log(AGE)$	0.51	3.17
Specific constants		
$SEXM$	0.29	2.75
P_1	0.69	4.81
P_2	-0.03	0.23
P_3	-0.09	0.82
P_4	-0.14	0.88
P_5	-0.1	0.66
P_6	-0.68	4.75
P_7	-0.55	3.7
P_8	0.32	1.52
P_9	1.12	10.59
P_{10}	-0.12	1.11
P_{11}	-0.96	2.34
P_{12}	---	---
GR_I	0.22	1.19
GR_{II}	-0.004	0.01
GR_{III}	0.11	0.68
GR_{IV}	---	---
GR_V	0.24	1.69

Linear expenditure function recommended by RETINA

Sample size	614	
Adjusted R^2	0.779	
Standard error of estimation	0.647	
RCMSPE	0.689	
AIC	-0.813	
Variable	**Coefficient**	***t*statistic**
Constant	7.87	5.65
Interactions		
$1/(\log(AGE))^2$	-29.28	3.7
$(\log(AGE))^2$	-0.19	3.24
$1/\log(IS)$	-0.12	1.59
Specific slopes		
$GR_{II}/\log(AGE)$	-3.73	5.53
$P_1/\log(AGE)$	0.72	2.71
$P_1/[\log(AGE)]^2$	-0.12	1.58
$P_2/[\log(AGE)*\log(IS)]$	4.4	2.77
$P_2/[\log(AGE)]^2$	-0.2	2.37
$P_2*[\log(AGE)/\log(IS)]$	-0.43	2.46
$P_2*\log(IS)$	-9.54	4.35
$P_2*[\log(IS)]^2$	-6.94	4.33
$P_3/\log(AGE)$	3.04	13.63
$P_3*\log(AGE)*\log(IS)$	-0.45	2.01
$P_4*[\log(AGE)/\log(IS)]$	-0.11	1.46
$P_4/[\log(AGE)]^2$	-0.07	0.9
$P_5*\log(AGE)*\log(IS)$	-0.67	2.91
$P_5/[\log(AGE)]^2$	7.85	2.66
$P_5*[\log(IS)/\log(AGE)]$	6.04	1.87
$P_6*\log(AGE)*\log(IS)$	-0.21	3.36
$P_6/[\log(AGE)*\log(IS)]$	-0.87	3.98
$P_7/\log(AGE)$	2.59	2.57
$P_7*\log(AGE)*\log(IS)$	-0.14	1.84
$P_7/[\log(AGE)*\log(IS)]$	1.47	1.23
$P_8*[\log(AGE)]^2$	0.25	1.97
$P_8/[\log(AGE)]^2$	100.78	1.92
$P_8/\log(AGE)$	-37.06	1.79
$P_9*[\log(AGE)/\log(IS)]$	-0.07	5.46
$P_9*\log(IS)$	-0.95	6.21
$P_{10}/[\log(AGE)*\log(IS)]$	-3.15	2.75
$P_{10}/\log(IS)$	0.9	2.49
$P_{10}*[\log(AGE)]^2$	0.08	4.75
$P_{11}*[\log(AGE)]^2$	0.05	2.63