E-commerce and digital divide in Spain using individual panel data 2008-2016.

04/2019

Preprint

Teodosio Pérez-Amaral *†

Angel Valarezo *†

Rafael López *†

Teresa Garín-Muñoz ‡

Iñigo Herguera *†

* Instituto Complutense de Análisis Económico, ICAE, Spain

† Universidad Complutense de Madrid, UCM, Spain

‡ Universidad Nacional de Educación a Distancia, UNED, Spain

Keywords: e-commerce; digital divide; TIC-H survey; individual annual panel data; logistic regression; dynamic model; Heckman selection model.

JEL Classification: C25; D11; O33

Acknowledgements: Presented at the ITS European Conference, 2018, Trento. Research funded by the Autonomous Community of Madrid, Spain. Project: Finance, Innovation and Strategies: Economic Analysis of Business Productivity and its Determinants (PRODECON-CM HM S2015/HUM-3491), 2016-2018.

ABSTRACT

Digital technologies sometimes create digital divides. One of the remedies for certain divides in Europe is the creation of the Digital Single Market, of which e-commerce is one of the main elements. The focus of this work is e-commerce in Spain. The current study improves substantially on existing international literature by using a large and representative panel data set on individuals, with 133,420 observations for the period 2008-2016. Moreover, it uses economic models that incorporate previously omitted variables, and employs a variety of panel techniques. This paper starts by measuring digital divides and their evolution along time. Next, a model that incorporates previously neglected explanatory variables, such as income and digital skills, is formulated. Individual demand models are estimated using panel logistic regression techniques. This allows quantifying the impact of each of the socioeconomic and geographic characteristics on the adoption of the service. The resulting models have high explanatory power. Newly incorporated variables like income and digital skills are highly significant. Age, education, gender and geographical variables are also significant. The results also allow novel regional comparisons. Policy recommendations are derived, suggesting effective and affordable measures targeted at specific socio-demographic groups.

1. INTRODUCTION

The Internet is one of the most powerful agents of change in recent history. Since the world wide web opened up the Internet for everyone in 1991[1], its diffusion has been so rapid around the globe and across new services that it has given rise to major changes in everyday life. According to recent data provided by the International Telecommunications Union (ITU, 2016) almost half of the world's population is using the Internet. However, there are important differences according to countries and groups of individuals.

E-commerce is an important service of Internet and a European Union priority because it has social benefits, promotes competition and advances towards the objective of a digital single market while avoiding digital exclusion[2]. Another reason for focusing on e-commerce is the relatively low penetration in Spain versus several European Union countries. The main goals of this study are to analyze and promote the usage of e-commerce.

Specifically, e-commerce has had important economic and social implications. Customers are changing their consumption habits and businesses need to adapt to the new situation modifying their business and marketing strategies. This situation is evolving rapidly, and it could be considered a sort of social revolution. At this point some authors wonder whether conventional shopping centers will survive or are doomed to disappear. In the early days of e-commerce, many consumers searched for information and compared products on the Internet, but they ultimately went to the physical store to shop. Now in some cases the situation is the opposite: some individuals use the physical store as a catalog where to feel and try the products that may then be bought online[3].

This paper analyzes e-commerce using microdata from the "Survey on Equipment and Use of ICT in Households" (INE, 2017) which is a survey similar to those carried out in the European Union and allows interesting comparisons across the 28 countries. Dynamic logistic models of adoption of e-commerce in Spain are estimated for the first time using individual panel data.

E-commerce has been an important subject of study as well as a political priority in the European Union and other international institutions; see for instance OECD (2001). There is a great amount of literature on the topic, which includes international and national studies. Producing a comprehensive survey of this literature is beyond the scope of this research. Previous studies are Cerno and Pérez-Amaral (2006, 2009), Garín-Muñoz and Pérez-Amaral (2011), Garín-Muñoz et al. (2018) and Valarezo et al. (2018). Moreover, references are given below regarding individual papers that share specific elements with the current work.

[1] There is broad consensus in situating the birth of the World Wide Web in 1991 when Tim Berners Lee created it, combining three existing resources (HTML ─Hypertext Markup Language, HTTP ─Hypertext Transfer Protocol, URL ─Uniform Resource Locator) with a new program called Web Browser.

[2] It is worth pointing out that digital exclusion refers to a variety of Internet services, not specifically to e-commerce.

[3] Nevertheless in Spain there is a high density of shops that are close to the final user, so this may be a partial explanation for the lower e-commerce figures.

A report by the National Telecommunications and Information Administration (1999) discusses the digital divides[4] (telephone, computer and Internet use) and their evolution in the US using data for the previous 15 years. They define implicitly the concept of digital divide[5] (absolute and relative) and they use it together with its evolution along time. The way of interpreting the level and evolution of the digital divide is similar to the one in this paper.

Limayem et al. (2000) study e-commerce empirically. They use the Theory of Planned Behavior with a longitudinal sample of 705 consumers from the US who were contacted by e-mail. Their methodology is not directly comparable with that of the present work. Theirs is based on dealing mostly with unobservable variables, while ours deals with observable variables.

Demoussis and Giannakopoulos (2006) study the determination and extent of Internet use as a facet of the digital divide in Europe. They use 2002-2003 individual cross section data from a variety of European countries to analyze the decision to use or not use the Internet as well as how much to use it. This paper contributes very useful discussions on the right way to account for the availability of Internet and the relevant population groups concerned. This insightful study, however, is missing a variable to measure digital skills and the estimates may also have biases due to the unobserved individual heterogeneity that has not been dealt with.

Pérez-Hernández, J. and R. Sánchez-Mangas (2011) analyze online shopping jointly with having Internet at home using the ICT-H survey of INE in Spain for the period 2004-2009 employing pooled data. In this paper they cannot control for unobserved individual heterogeneity since they use pooled instead of panel data. Likewise, they omit relevant variables like digital skills, geographic variables and individual income (which is available only from 2008 onwards). This can cause inconsistency in the estimates as recognized by the authors on p. 221.

Hilbert (2011) analyzes the digital divides and their definitions. The author considers the adoption of a new service as a contagious disease instead of an economic decision that would depend mostly on income, price and other factors. This paper suggests that there are heterogeneous digital divides which cannot be added up in a significant way, since they are difficult to synthesize in a single index. This paper poses relevant questions that would need specific answers.

Robles and Torres-Albero (2012) analyze individual data from INE for the year 2009 to calculate penetrations of the use of Internet. They conclude that "…between the most advanced communities and the communities with the lowest percentage of users these differences, far from decreasing, have remained

[4] "The gap between individuals, household, business and geographic areas at different socio-economic levels with regard to their opportunities to access information and communications technologies and to their use for a wide variety of activities" (OECD, 2001, pp. 8-9).

[5] A digital divide can be regarded as a situation in which the demand for access or use of a given digital service by a specific group is considered as insufficient (by policy makers or researchers), compared to that of other group of reference.

stagnant or even increased slightly over the last five years". They also model the decision of using the Internet. Their logistic model does not control for individual heterogeneity and fails to include two central variables such a digital skills and income that were available at the time. This leads to inconsistent estimates and an upward bias in the estimation of the effect of "level of education" which is positively correlated with the two omitted variables. These flaws undermine some key conclusions of the paper.

Hackl el al. (2014) analyzes the interaction between market structure and market performance of e-commerce centered on supply side considerations of the retail sector for the case of Austria, using cross section data.

A report by Correa et al. (2015) of BBVA Research analyzes previous waves of the same TIC-H survey of INE for Spain during the period 2008-2014. They analyze the adoption of broadband, and e-commerce. They use pool data, not a panel, so they cannot control for unobserved individual heterogeneity. They do not use an explicit economic model, but they do use time dummies and geographical variables. Their conclusions are subject to caveats due to the type of sample and model they use and the importance that they assign to e-banking.

The European Commission (2018) shows that Spain has an average level of digital development within Europe. The main barriers for development are in the demand side, in particular the shortcomings in human capital related to ICT as consumers. The document contains a useful account of recent and current programs for the advancement of the Information Society in Spain, both from the supply and the demand sides.

A recent report by Fundación BBVA (2018) found that the digital divide in Spain disappeared in 2017. However, they only analyze, at an aggregate level, the digital divide in access and use of Internet but not e-commerce, like the present study does, which is more relevant for the purposes of this paper. They use descriptive statistics as well as intuition, but they do not use economic or econometric models. The effect of mobile broadband access is not explicitly considered.

Another study by Burgos et al. (2018) is centered on e-commerce in Spain. It uses a highly mathematical model based on epidemiology but fails to consider the economic nature of the decision of adopting e-commerce as well as the findings in previous literature. The treatment of the data seems perfunctory and is not sufficiently explained. The multiple factors that may affect the adoption of e-commerce are largely ignored.

A recent study by Răileanu (2018) analyzes the regional digital divide in the European Union. It uses panel data specific models aggregated at the levels of regions and countries, but no formal theoretical model is used. Variables such as digital skills are absent, and the level of education turns out to be insignificant in some models. The endogeneity of other explanatory variables may be an issue. The interpretation of the signs of the coefficients of several variables seems to be problematic in several cases.

The current paper improves substantially on existing literature by using a 9-year-individual panel, providing disaggregate measures of digital divide, and using an explicit and richer economic model for individual panel data. Additional relevant variables, such as income, digital skills and geographical variables are included as well as a variety of panel estimation techniques. The focus here is on specific individual level policy recommendations. The results reported may be amenable to further exploitation by other researchers.

The rest of the paper is organized as follows: Section 2 contains an overview of e-commerce in Spain as well as its recent evolution. In Section 3 the data is presented, highlighting the construction of the panel database on individuals from the original data on dwellings. Section 4 presents the economic models showing the variables that influence the probability that an individual becomes an online consumer. The main conclusions, policy recommendations, caveats, and further research are shown in section 5.

2. OVERVIEW

Figure 1
Internet and e-commerce use as percentage of people aged 16 to 74 in Spain (2008-2017)

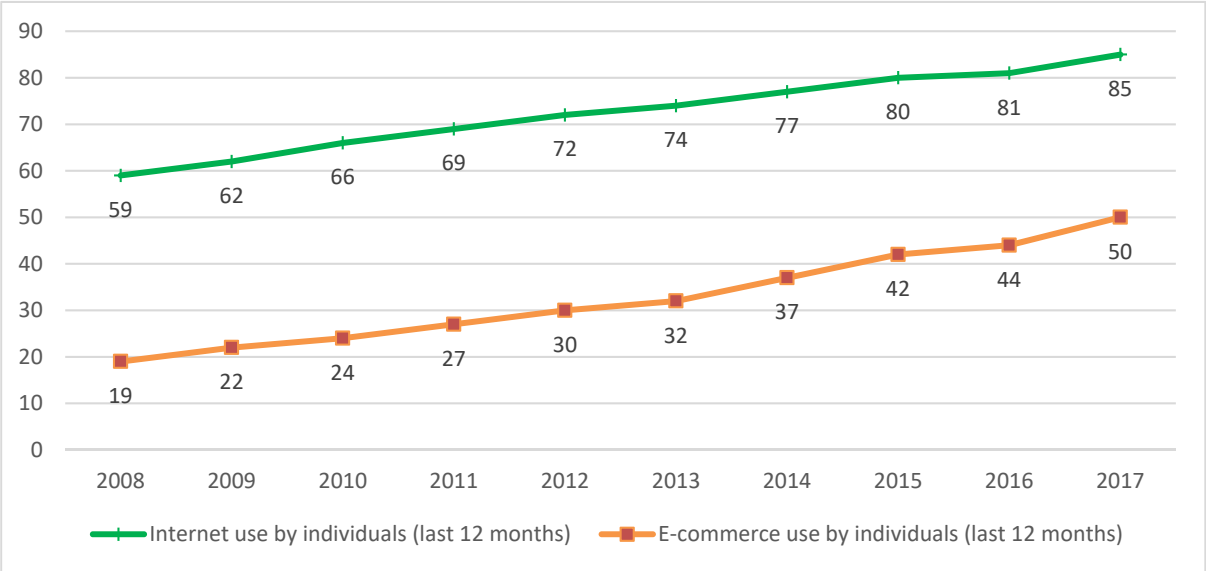


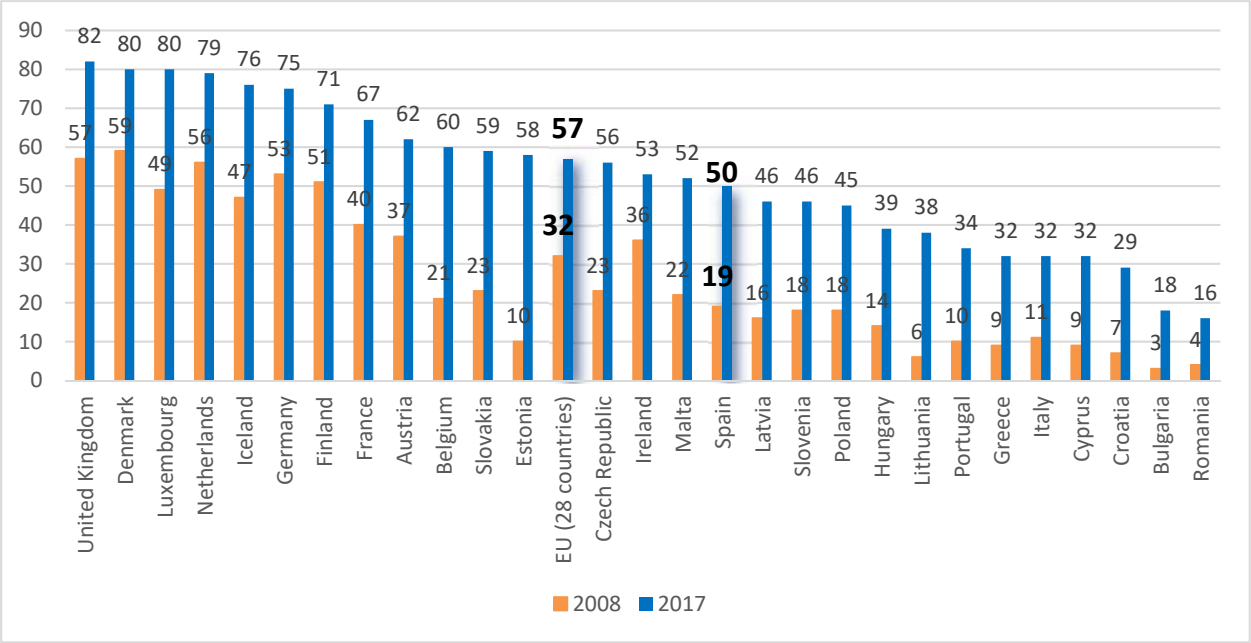
Figure 1 shows the evolution of the aggregate use of Internet and e-commerce in Spain for the years 2008 to 2017, obtained from Eurostat (2017). The penetration of both services grows over time. The percentage of Internet users increases from 59% to 85% of the population aged 16 to 74. Likewise, the percentage of population that uses e-commerce grew from 19% to 50% of those aged 16 to 74[6], which suggests that there has been a strong increase, but that there is still room for improvement.

Figure 2 shows an international comparison of the penetration rates of e-commerce for European Union countries and for Spain in 2008 and 2017. The European Union average increases from 32% to 57% in the period, while Spain increases from 19% to 50%. The increase in penetration is 31pp in Spain and 25 pp for

[6] The data of our survey, INE (2017), contains information on people of all ages, including those over 74 years of age. However the data in this section contains information only on people up to the age of 74, for compatibility with the corresponding data of Eurostat.

the European Union between 2008 and 2017. The gap is closing significantly between Spain and the European average by 6pp in the period considered 2008-2017. However, there is still a significant gap for the first seven countries on the left of figure 2, which are those that could be the reference for the policy objectives of Spain. This gap is still around 28 pp in 2017, when comparing it with the average of the seven countries.

Figure 2
Penetration rates of e-commerce in the EU-28 as percentage of people aged 16 to 74 (2008 and 2017)



3. THE MICRODATA

a. Treatment of the sample data.

This paper starts by analyzing annual data on individual dwellings from the survey on Equipment and Use of Information and Communication Technologies in Households (ICT-H Survey) from 2008 to 2017.

The survey is conducted by the Spanish National Statistical Institute (INE, 2017) using the methodology of Eurostat. It includes an elevation factor, and it is representative at both national and regional levels. The raw data are available at INE (2017), as well as a variety of documents related to the design, implementation and statistics (except for the individual identifier). It is a rotating survey which includes 15,000-20,000 dwellings each year. The same dwelling is interviewed a maximum of four (consecutive) years, and around 30% of the dwellings are replaced every year. 60% of the interviews are conducted by phone (CATI) and 40% in person (CAPI). Following Eurostat's guidelines, similar surveys have also been performed in the rest of the EU countries.

During the writing of this paper, Pool data were available for 2008-2017, but Panel data could only be used for the period 2008-2016 due to the lack of a dwelling identifier for 2017. The raw panel data underwent

meticulous analysis and filtering to extract and homogenize the information about individuals uniquely identified throughout the observation period. The process goes as follows:

Pool data

Read the information on dwellings for each year (publicly available from INE, see INE 2017).

Homogenization of variables across years (using our own custom-made algorithm)

Panel data

Incorporate the dwelling identifier supplied by INE to the previous Pool database.

Apply our custom-designed filters to extract a panel of distinct individuals from the Pool of dwellings.

Cleaning and treatment of outliers.

At the end we obtained a Panel 2008-2016 database consisting of 133,420 observations (corresponding to 59,252 different individuals) and around 700 different variables.

This was the final micro Panel database that was used as in all of our models.

b. The explanatory variables used are classified as follows:

Sociodemographic	
Gender:	2 groups: 1 if male, 0 if female
Age:	6 groups
Habitat:	4 groups
Household Members:	5 groups
Nationality:	2 groups: Spanish, Foreign
Individual skills	
Education:	4 levels of study
Digital Skills:	4 levels
Economic	
Employment Situation:	6 groups
Income:	4 groups, monthly net income
Time and Geographic	
Yearly Dummies:	1 for each year
Regional Dummies:	17 Autonomous Communities + 2 Autonomous cities

The variables were categorized as shown in table 1, which also contains the penetrations of e-commerce on the different groups, measured in the years 2008 and 2017. The table suggests a general increase in the penetrations, which is compatible with convergence towards European standards at an aggregate level.

Table1.
Penetration rates of e-commerce in Spain (2008 and 2017)

Characteristics	Categories	2008	2017
Gender	Female	14.2	41.4
	Male	20.5	48.5
Age	<25	25.1	64.5
	[25,35)	31.4	68.1
	[35,45)	21.3	62.3
	[45,55)	16.2	51.0
	[55,65)	7.2	31.8
	65+	1.0	9.3
Education	Primary	1.4	6.5
	Secondary	16.6	44.7
	Bachelor's Degree	29.7	71.1
	Master´s or PhD	42.6	81.8
Digital Skills	Low	9.8	12.3
	Medium	25.1	44.5
	High	47.4	71.2
	Very High	69.2	89.8
Habitat	500,000+	22.2	51.2
	100,000-500,000	18.2	42.2
	20,000-100,000	16.3	45.4
	<20,000	13.0	38.7
Household Members	1	12.4	31.1
	2	14.6	32.5
	3	18.1	51.2
	4	21.5	58.7
	5+	14.9	43.0
Employment Situation	Employed	25.1	63.9
	Unemployed	12.7	38.8
	Retired	2.2	13.3
	Student	27.8	64.2
	Housekeeper	4.0	13.1
	Other	11.2	27.6

Note: percentage of individuals between 16 and up that have used ecommerce for private purposes at least once in the previous 12 months.

Table 1 (continued).
Penetration rates of e-commerce in Spain (2008 and 2017)

Characteristics	Categories	2008	2017
Nationality	Foreigner	14.5	41.0
	Spanish	17.7	45.3
Income	Low	3.6	23.6
	Medium	11.8	37.5
	High	28.3	61.0
	Very High	41.0	79.7
Autonomous Community or Autonomous City	Andalucía	13.1	41.0
	Aragón	18.7	49.1
	Asturias	16.1	44.3
	Baleares	24.3	52.9
	Canarias	12.2	37.6
	Cantabria	18.6	49.8
	Castilla y León	14.6	39.4
	Castilla-La Mancha	12.6	41.7
	Cataluña	21.2	49.1
	Valencia	15.3	43.7
	Extremadura	14.4	37.1
	Galicia	13.9	36.0
	Madrid	23.8	53.9
	Murcia	14.2	39.2
	Navarra	20.1	53.2
	País Vasco	21.1	46.7
	La Rioja	18.3	44.0
	Ceuta	12.6	48.6
	Melilla	17.9	47.8
TOTAL		17.3	44.9

Note: percentage of individuals between 16 and up that have used ecommerce for private purposes at least once in the previous 12 months.

Graphs of Figures 3 and 3.1 selected entries of table 1 and the values of the relative digital divides. It shows the evolutions of the penetration of e-commerce by selected categories, also the absolute divide in perpendicular. It is worth pointing out that the absolute digital divides for gender, age, education and income increase along time. They are measured in terms of the absolute difference between the two rates of penetration in 2008 and 20177. The absolute age divide increases by 31.1 pp, while the gender divide increases by 0.8 pp, the education divide increases by 34.1 pp, the income divide increases by 18.7 pp and the employment divide increases by 29 pp along the 10-year period.

[7] An alternative measure of the relative divide consists in dividing the difference between the two categories at a given point in time over the highest of the two rates of penetration. This measure can give different results from the absolute measure. An additional measure of divide can be the ratio of penetration rates. That is, the ratio between two penetration rates of two different categories at a given point in time, using the highest as the denominator. Different measures may generate apparent contradictions and different impressions for the readers. Graphs like those in figure 3 may be useful to help the researcher or policy maker decide for him/herself.

However, the relative divides measures shown (gender, age, education, income and employment) decrease by 16.1, 10.4, 4.6, 20.8 and 3.6% respectively.

These variations in inequality may be partly due to the economic crisis that started in 2007. They also may be due to the tendency that at early stages of a given service the digital divides tend to increase with time while at a later stage, when the market of a service is more mature, the digital divides tend to disappear. There are income effects, demographic effects and diffusion effects that may work in the direction of closing the gaps that may appear in the short run. This is left for further research.

OECD (2001) deals specifically with the digital divides, although the report does not provide a precise definition of digital divide, or a method to compare digital divides at two different points in time, or its temporal evolution. An additional reference for the specific case of Spain is Robles and Torres (2012).

Figure 3.
Digital divides: penetration rates by Gender, Age, Education and Income in Spain 2008-2017

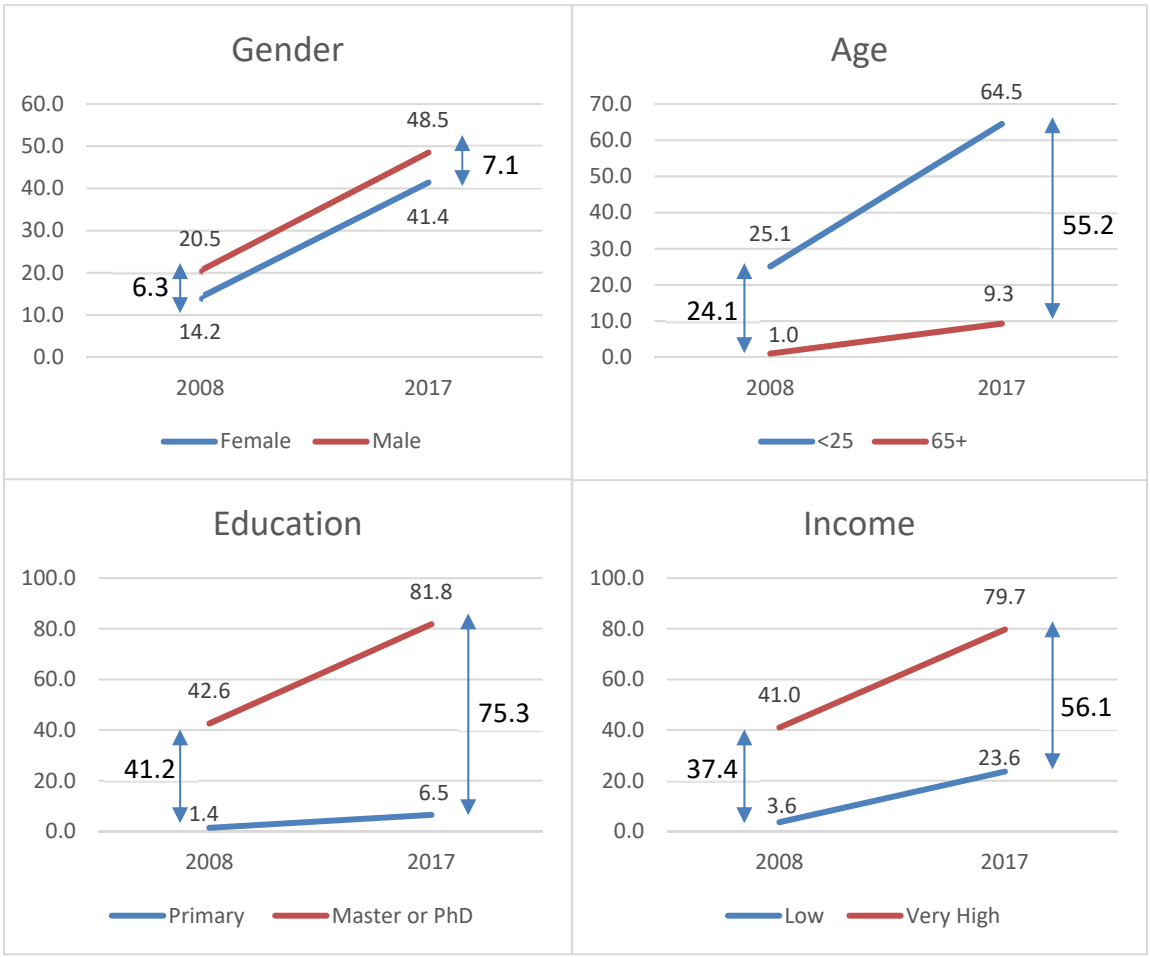


Figure 3.1.
Evolution of penetration rates by Employment Situation in Spain 2008-2017

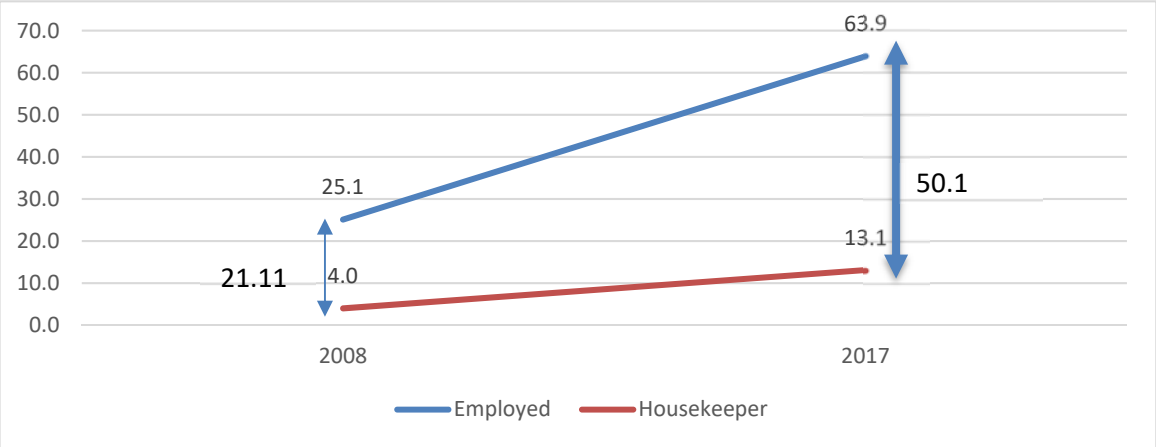
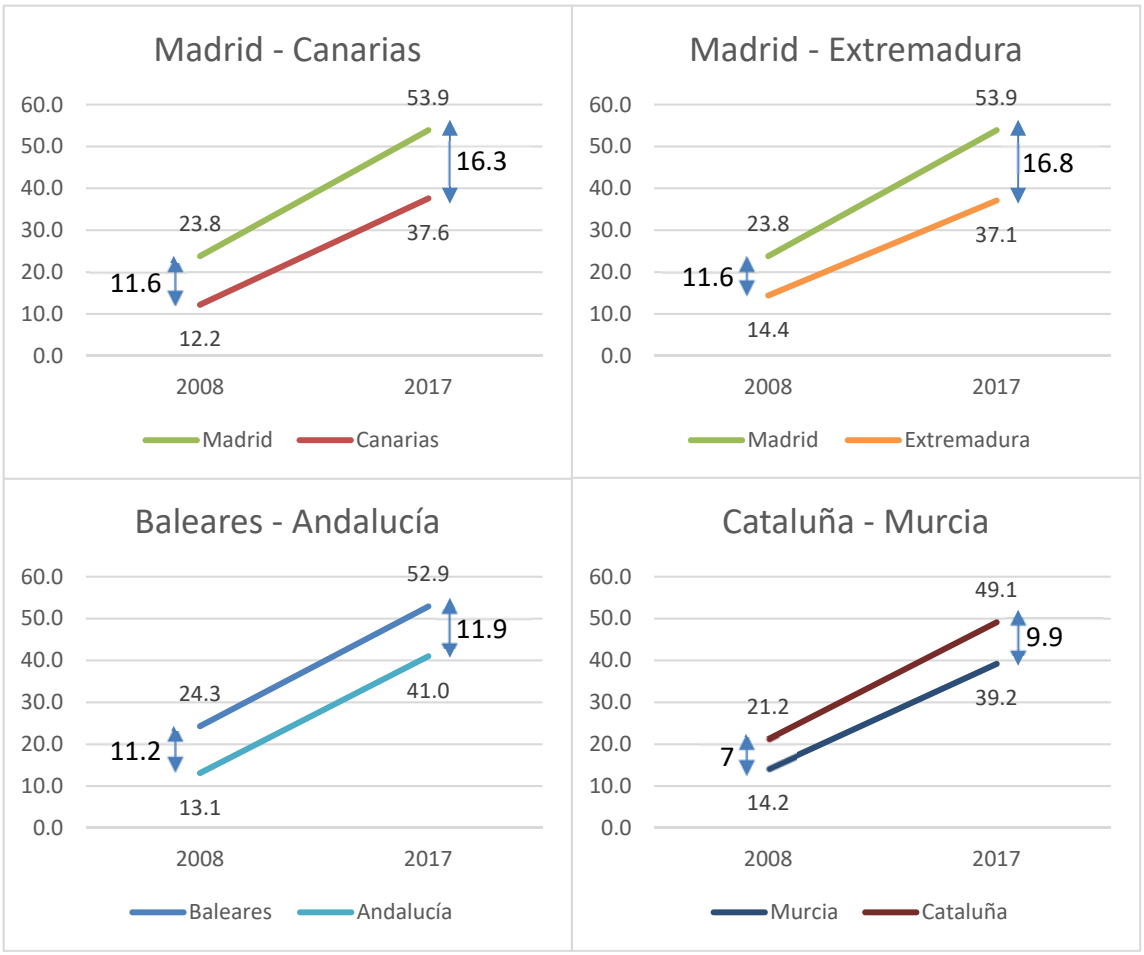


Figure 4
Evolution of selected penetrations of e-commerce by regions in Spain 2008-2017



A similar situation occurs when the regional differences are analyzed. The digital divides between selected regions are shown in figure 4, where the digital gaps between richer and poorer territories seem to be increasing by 4.7, 5.2, 0.7 and 2.9 pp respectively in the specific regions shown. Meanwhile the measures

of relative gap decrease along time. Notice that the choice of cases shown is not random. The cases that display less convergence are chosen here to underscore the overall convergence of the different groups and regions along time.

An exercise similar to this one using data from services other than e-commerce and for each European country is feasible but beyond the scope of this paper. The slow convergence of some of these divides may be a sign of short-term e-exclusion. This may be the rationale for targeted policy interventions at the national, regional or individual levels. We leave this for further research.

4. MODELS AND EMPIRICAL RESULTS

The present study follows an economic perspective using the neoclassical utility maximization approach, (Varian, 2002). The demand for access is determined by the size of consumer surplus associated with Internet usage and the cost of access. Regarding access to e-commerce, the relevant theory is that of the telephone demand framework of Artle and Averous (1973), Squire (1973), Von Rabenau and Stahl (1974), Rohlfs (1974), Taylor (1994), Kridel et al. (1999) and Rappoport et al. (2002).

In telecommunications the use of a service is conditional on the access to this service, (Taylor (1994)). An assumption of the current approach is that Internet access is a prerequisite for adopting e-commerce. In any case, Internet access could be obtained in the period considered, through a variety of channels and places: buses, trains, airports, ships, work, home, school, university, hotels, restaurants, public Wi-Fi zones, community access centers, libraries, post offices, internet parlors, as well as using a variety of technical solutions: dial-up, cable, ADSL, broadband, narrowband, or through mobile phones, tablets and portable computers. Summing up, Internet access has been ubiquitous during the years of the sample 2008-2016.

In many cases access to the Internet is not an explicit decision, but rather a circumstance governed by the commercial policies of carriers that incorporate Internet to a traditional service, even without explicit knowledge by the consumer. A similar argument is sustained by Demoussis and Giannakopoulos (2006) for the European case using 2002-2003 data. The argument is more compelling in the sample considered. Nevertheless, in the Appendix, table A1, a model incorporating a Heckman-style mechanism is estimated and presented, suggesting that the selection mechanism is not necessary in this case and model 1 is an adequate approximation.

Taking widespread access to the Internet for granted, the decision to use the Internet no longer needs to be modelled. The hypothesis is that consumers decide to use e-commerce given that they have access to the Internet[8].

[8] Effective use by 100% of individuals across the population cannot be expected, since there are people who are severely ill, physically or mentally handicapped, very old, very young, and minorities for whom Internet may not be attractive.

In this context, an individual derives utility (U) from adopting a particular Internet service (Y), if the benefits from using that service B(Y) exceed its costs C(Y). Empirical works based on this approach are Demoussis and Giannakopoulos (2006); Fairlie (2004); Vicente and López, (2008); Lera-López et al., (2011) and Valarezo et al. (2018, 2019), the last four referred to the case of Spain.

From a standard neoclassical utility optimization approach, the maximization of the utility (U) of an individual obtained from e-commerce (Y_i), will be a function of the benefits $B(Y_i|x)$ of doing so and the costs $C(Y_i|x)$, where x is a set of conditioning variables associated with it. The conditional probability of using e-commerce is:

$$P(Y_i|x_i) = P[B(Y_i) - C(Y_i) > 0|x_i] \quad i = 1, \dots, N, \ individuals. \qquad \textbf{(1)}$$

The adoption of e-commerce is modeled using a binary dependent variable for each of the estimated empirical models. The individuals considered are those e-commerce users who purchased online goods or services for private use in the last year with respect to the total population, in the first model.

This paper improves on previous literature on e-commerce by using a large and representative panel data base of individuals which allows for the treatment of individual heterogeneity and obtaining consistent estimates. A variety of different economic models is used for assessing robustness. Previously excluded variables, such as income and digital skills, are included. This allows for consistent estimation of their effects and those of related variables. An interpretation as an income effect and a cost effect in an economic model is natural in this context. By employing appropriate econometric models, including logistic regression, dynamic and also sample selection correction models, estimates that are consistent, efficient and robust are obtained.

The explanatory variables of the models have been introduced in the previous section. Lagged values of e-banking and e-government, which are meant to capture the dynamics and the cross effects on e-commerce are also introduced as explanatory variables in the second model.

Table 2 contains the two logistic specifications of the model, static and dynamic, including the point estimates of the odds ratios and the corresponding z statistics[9]. Odds ratios greater than one suggest a positive effect on the adoption of e-commerce relative to the reference group, while lower than one corresponding to a negative effect.

Both models are similar in several respects. The signs and sizes of most coefficients and z statistics are similar and have similar interpretations. This can be regarded as a sign of robustness of the models.

The two lagged variables, e-banking and e-government are significant with positive effects. However, there seems to be little gain in the dynamic model, which loses almost half of the observations due to the lags, so

[9] All estimations are performed using Stata 15.

we concentrate on the results for model 1. For the sake of brevity, only results that may have policy implications are discussed.

Table 2.
Models of adoption of e-commerce by individual Internet users.
Panel data (2008–2016)

	Variables		(1) E-commerce static logistic model		(2) E-commerce dynamic logistic model	
			Odds ratios	z	Odds ratios	z
Lagged variables	ebanking$_{t-1}$				**2.56**	**21.37**
	egovernment$_{t-1}$				**1.31**	**6.20**
Sociodemographic	Gender	Female				
		Male	**1.45**	**11.90**	**1.52**	**9.51**
	Age	16-24				
		25-34	0.85	-0.81	0.94	-0.16
		35-44	**0.56**	**-3.12**	0.49	-1.82
		45-54	**0.38**	**-5.03**	**0.38**	**-2.49**
		55-64	**0.23**	**-7.19**	**0.26**	**-3.30**
		65<	**0.14**	**-8.37**	**0.17**	**-4.00**
	Education	Primary or less				
		Secondary	**1.58**	**7.12**	**1.51**	**3.94**
		Bachelor	**2.34**	**11.95**	**1.98**	**5.99**
		Master/Phd	**3.36**	**16.86**	**2.68**	**8.64**
	Digital Skills	Low				
		Medium	**2.56**	**4.90**	**2.74**	**2.55**
		High	**7.95**	**11.22**	**7.16**	**5.17**
		Very high	**25.32**	**17.36**	**24.29**	**8.32**
	Habitat	<20,000				
		20,000-100,000	0.94	-1.49	1.00	-0.03
		100,000-500,000	**0.78**	**-3.82**	**0.73**	**-3.52**
		500,000<	1.00	-0.05	1.03	0.63
	Household Members	One				
		Two	**0.87**	**-2.69**	0.87	-1.78
		Three	**0.73**	**-6.02**	**0.75**	**-3.74**
		Four	**0.71**	**-6.42**	**0.76**	**-3.57**
		Five or more	**0.60**	**-7.15**	**0.64**	**-4.35**
	Nationality	Foreigner				
		Spanish	**1.26**	**3.67**	1.00	-0.02
Economic	Employment Situation	Employed				
		Unemployed	**0.73**	**-6.94**	**0.76**	**-4.21**
		Retired	**0.84**	**-2.16**	0.84	-1.52
		Student	**0.70**	**-4.78**	**0.77**	**-2.38**
		Housekeeper	**0.81**	**-2.79**	0.83	-1.66
		Other	0.83	-1.89	0.93	-0.51
	Income	Low				
		Medium	**1.60**	**10.43**	**1.52**	**6.01**
		Medium-high	**2.58**	**19.04**	**2.42**	**11.74**
		High	**4.00**	**23.43**	**3.67**	**14.73**

Table 2 (Continued)

	Variables		(1) E-commerce static logistic model		(2) E-commerce dynamic logistic model	
			Odds ratios	z	Odds ratios	z
Interaction	Digital Skills × Age	High × 55-64	**2.44**	**4.02**	1.62	1.15
		High × 65 or more	**2.33**	**3.24**	1.53	0.92
		Very high × 55-64	**1.80**	**2.53**	1.21	0.44
		Very high × 65<	**3.09**	**3.90**	1.67	1.05
Time effects	Year	2008				
		2009	**0.84**	**-3.28**	0.87	-1.85
		2010	**1.17**	**2.88**	1.11	1.28
		2011	1.05	0.76	0.91	-1.10
		2012	**1.13**	**2.03**	1.16	1.77
		2013	**1.23**	**3.57**	**1.20**	**2.22**
		2014	**2.09**	**12.57**	**2.01**	**8.25**
		2015	**4.21**	**24.33**	**3.90**	**15.94**
		2016	**4.09**	**23.94**	**3.90**	**16.04**
Sociogeographic effects	Autonomous Community	Andalucía				
		Aragón	1.16	1.81	1.25	1.96
		Asturias	**1.26**	**3.01**	**1.33**	**2.65**
		Baleares	**2.35**	**9.39**	**2.07**	**5.37**
		Canarias	**0.82**	**-2.24**	**0.69**	**-2.86**
		Cantabria	**1.60**	**5.36**	**1.67**	**4.11**
		Castilla la Mancha	**1.26**	**2.71**	**1.29**	**2.09**
		Castilla León	1.04	0.45	1.15	1.28
		Cataluña	**1.55**	**6.58**	**1.40**	**3.60**
		Extremadura	1.12	1.29	1.26	1.79
		Galicia	1.04	0.46	0.97	-0.32
		La Rioja	**1.27**	**2.58**	**1.36**	**2.35**
		Madrid	**1.42**	**5.16**	**1.32**	**2.93**
		Navarra	**1.66**	**6.92**	**1.85**	**5.83**
		País Vasco	**1.79**	**7.27**	**1.72**	**4.83**
		Murcia	0.94	-0.75	0.98	-0.18
		Valencia	0.91	-1.34	0.86	-1.57
		Ceuta	0.72	-1.72	0.73	-1.10
		Melilla	0.98	-0.11	0.87	-0.41
	Constant		0.05	-15.36	0.02	-10.25
	N. observations		66,169		34,032	
	Wald χ^2		7982.03, DF: 69		3704.33, DF: 71	

Notes: Odd ratios and z statistics significant at the 5% are represented in bold. Random-effects logistic model (Equation 1). Random effects dynamic logistic model (Equation 2). Heteroskedasticity robust estimates. Estimations refer to all Internet users that have purchased (or not) online in the last year. The first category of each variable is the reference, and it is omitted. The samples reflect the fact that some variables are missing for some observations, which reduces the usable sample size. The dynamic model loses almost half of the sample due to the lagged variables. All estimations have been calculated using Stata 15.

The estimates are presented in the form of odds ratios; estimated odds ratios with values smaller than one are associated with variables that have a negative effect on the dependent variable. Odd ratios above one

are associated with explanatory variables that have a positive effect on the dependent variable. Individual significance of each coefficient is tested using the z (standard normal) statistic. Corresponding z statistics are calculated from the null of the odd ratio being equal to one (no effect). Odds ratios below one have corresponding negative values of z, while odds ratios above one have corresponding z with positive signs.

The interpretation of the estimation results is as follows: Males have higher odds of using e-commerce, while age has a negative impact on e-commerce for people with low levels of digital skills, as shown in figure 5 of the Appendix. Education and digital skills have important positive effects, while the effect of habitat is somewhat negative for populations between 100,000 and 500,000 inhabitants.

The number of household members has a negative effect, while being Spanish has a positive effect. Those employed tend to do more e-commerce, while income has a positive and strong effect. Notice that the two variables that have been mostly neglected in previous literature, income and digital skills, are both highly significant here, which suggests that we have more accurate estimates that are more reliable than in the past.

The time effects, shown in figure 6 of the Appendix are dummies that can be thought of as capturing other demand side variables, like time varying effects and the economic crisis, as well as supply side effects not included in the explicit explanatory variables. The time effects are significant and mostly growing over time.

The geographical characteristics are captured by the dummy variables corresponding to the 17 autonomous communities (and two autonomous cities, namely Ceuta and Melilla). They capture geographic specific effects not accounted for in the rest of the model. They would be essentially autonomous community individual effects. In our case it is apparent that communities that are richer tend to do more e-commerce, like Baleares, Cataluña, Madrid, Navarra and País Vasco. They are joined by Cantabria that has made successful efforts to promote the information society.

In contrast, not so rich communities like Canarias, Murcia, Valencia and the city of Melilla tend to do less e-commerce. This variable seems to be capturing a geographical digital divide which is based on (private and public) income and wealth divide. In some cases, like the Canarias archipelago, the costs of delivery may be so high as to substantially limit the supply of certain products that are available in continental Spain.

Multicollinearity may be a concern in this model. However, due to the panel structure of the data together with the type of categorical data which are orthogonal within each variable, the issue may not be so relevant. In any case, polychoric correlations among selected explanatory variables have been computed to obtain formal tests, shown in table 3.

Table 3. Polychoric correlations among selected independent variables

	Gender	Age	Education	Digital Skills	Household Members	Income
Gender	1.00					
Age	0.02	1.00				
Education	-0.08	0.00	1.00			
Digital Skills	0.07	-0.29	0.43	1.00		
Household Members	0.02	-0.31	-0.02	0.06	1.00	
Income	0.07	0.06	0.46	0.30	0.24	1.00

The values of all off diagonal elements that lie below .5 suggest that there is a limited amount of multicollinearity between the data of the independent variables.

5. CONCLUSIONS

This paper improves substantially on previous literature on e-commerce by using a large and representative panel data base of individuals and by using a variety of economic models. Previously excluded variables, such as income and digital skills, are included allowing for a natural interpretation as an income effect and a cost effect in an economic model. By employing appropriate econometric models, including logistic regression, dynamic and also sample selection correction models, estimates that are consistent, efficient and robust are obtained.

Firstly, by analyzing the penetrations across different groups of individuals it can be noted that several digital divides appear for gender, age, education, digital skills, occupation and income, as seen in table 1. Some absolute digital divides do not tend to close over time while others do, however relative digital divides generally tend to decrease over time.

For isolating the effect of each of the determinants of the adoption of e-commerce, appropriate panel data techniques controlling for individual heterogeneity, time and geographic effects are used. The coefficients of the determinants for adoption are large and significant, in particular: gender, income, age, digital skills and education. Income has been included in the present model as is usual in a demand equation. This allows improved estimation and inference.

High digital skills have a positive influence on the adoption of e-commerce, partly counteracting the negative effect of some age groups. Insufficient digital skills seem to be partly responsible for the digital divide, as is highlighted by the European Commission (2018). This variable has been largely omitted in previous empirical literature on digital divide, leading to flawed inference.

Based on these conclusions some policy recommendations can be formulated as follows: when promoting e-commerce is a priority, several measures can be implemented at national and regional levels with the restriction of representing low costs on the part of the governments, firms and citizens. Most measures

proposed here are low cost, specific, and targeted at individuals or groups, independently of their geographical location.

Focusing first on demand side measures, for example, may be desirable to reinforce a training program on specific digital demand side factors, like digital skills (pointed out by the EC as a critical bottleneck for demand) in order to bridge the digital divides. These measures could be focused on females, people over 55, those with low digital skills and homemakers. A complementary measure could be to provide technical support online, by phone or in person, to those groups that are more at a disadvantage. A training program may be more effective when focused on those that could increase from low to medium level of digital skills.

Focusing on supply side measures, the government has recently implemented some interesting general measures, like the "digital by default" program (Estrategia 2015-2020), aimed at digitalizing all interactions of public services. It also created Offices of assistance on the use of public services for serving citizens using e-government, see European Commission (2017). Another measure for 2018 is the Grant program that offers basic training in ICT to young digital professionals so that they can gain access to jobs in this sector, see European Commission (2018).

Additionally, central and regional governments could promote complementary services such e-health and identify and re-edit supply side programs that were successfully implemented whether in their own territory or in others. The administrations could also reduce transaction costs, red tape and trade barriers. Some territories, like the archipelago of Canarias, still do not have access to some e-commerce transactions that are available in the mainland, possibly due to its long distance across the Atlantic Ocean and high logistic costs. Guidelines or incentives for efficient e-commerce platforms could also be established.

With respect to the private sector, practical recommendations could be to implement easy-to-use platforms, facilitate legitimate customer reviews for their products and services, and assure more security for payment and transaction processes (which includes the handling of credit card fraud).

These results and conclusions present some caveats and limitations. The data set is large, but it is declarative, not actual data, which limits the quality of the data due to the difficulty of recollection. Moreover, the data are not specifically designed for this research on e-commerce, which also limits the applicability of the results.

In terms of future research, the continuation of the analysis of the data set for topics like spending in e-commerce, broadband access, cloud services, trans-border e-commerce, use by children, etc. constitute a rich research agenda. An in-depth study of digital divide using individual data and employing additional indicators is another priority. Clearly stating its definition and formulas for its computation, as well as the calculation of its differences across groups and the interpretation of its evolution over time, is also in the research agenda. Testing the hypotheses that the evolution of a given digital divide is analog to a life cycle model along time is another avenue of research. On the other hand, the availability of data on actual behavior would be useful for defining new priorities of research.

REFERENCES

Artle, R. and Averous, C. (1973). "The telephone system as a public good: Static and Dynamic Aspects", Bell Journal of Economic and Management Science, Vol. 4, No 1, Spring, 89-100.

Burgos, C, Cortés, J-C, Shaikhet, L, and R-J Villanueva (2018). "A nonlinear dynamic age-structured model of e-commerce in spain: Stability analysis of the equilibrium by delay and stochastic perturbations." Commun Nonlinear Sci Numer Simulat 64, 149-158.

Calzada, J., García-Mariñoso, B., Ribé, J., Rubio-Campillo, R. and D. Suárez (2017). "Fiber deployment in Spain". UB Economics Working Papers 2017/364.

Cerno, L. and T. Pérez-Amaral (2006). "Demand for Internet Access and use in Spain", in PREISSI, B. and MULLER, J. (eds.), *Governance of Communication Networks, connecting societies and markets with IT*, Physica Verlag, Heildeberg.

Cerno, L., and T. Pérez-Amaral. (2009). E-Commerce Use in Spain. In P. Curwen, J. Haucap, & B. Preissl (Eds.), *Telecommunication Markets: Drivers and Impediments* (157–172). Heidelberg: Physica-Verlag HD. https://doi.org/10.1007/978-3-7908-2082-9_9

Correa, M, García, JR, and A. Tabanera (2015) "Comercio electrónico y hábitos de consumo en España: la importancia de la banca on-line" BBVA Research. Observatorio de la economía digital.

Demoussis, M. and N. Giannakopoulos (2006). "Facets of the digital divide in Europe: Determination and extent of internet use". *Economics of Innovation and New Technology*, 15, 235-246.

European Commission (2017) Digital Economy and Society Index Country profile (DESI) 2017: Profile of Spain, https://ec.europa.eu/digital-single-market/en/desi

European Commission (2018) Digital Economy and Society Index Country profile (DESI) 2018: Profile of Spain, https://ec.europa.eu/digital-single-market/en/desi

Eurostat (2017). ICT usage in households and by individuals from Eurostat http://ec.europa.eu/eurostat/cache/metadata/en/isoc_i_esms.htm

Fairlie, R.W. (2004). "Race and the digital divide, Contributions to Economic Analysis and Policy". *The Berkeley Electronic Journals*, vol. 3, 1-38.

Fundación BBVA (2018) "Uso de Internet en España y la Unión Europea" Esenciales IVIE 29/2018.

Garín-Muñoz, T. and T. Pérez-Amaral (2011). "Factores determinantes del comercio electrónico en España". *Boletín económico de ICE, Información Comercial Española*, vol. (3016), 51-65.

Garín-Muñoz T., López, R., Pérez-Amaral T., Herguera. I. and Á. Valarezo, (2019). "Models for individual adoption of eCommerce, eBanking and eGovernment in Spain", Telecommunications Policy Vol. 43, 100-111.

Hackl, F., Kummer, ME, Winter-Ebmer, R. and C. Zulehner, (2014) "Market structure and market performance in E-commerce" European Economic Review, 68, 199-218.

Hilbert, M. (2011) "The end justifies the definition: The manifold outlooks on the digital divide and their practical usefulness for policy-making"

INE (2017). "Survey on Equipment and Use of Information and Communication Technologies in Households". Telecommunications Policy 35, 715-736.

https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176741&menu=ultiDatos&idp=1254735976608

International Telecommunications Union, ITU (2016). ICT Facts and Figures.

Lera-Lopez, F., M. Billon and M. Gil (2011). "Determinants of internet use in Spain". *Economics of Innovation and New Technology*, 20(2), 127–152.

Limayem, M. Khalifa, M. and A. Frini (2000) "What makes Consumers Buy from Internet? A longitudinal Study of Online Shopping", IEEE Transactions on Systems, man and Cybernetics 30, No. 4.

Littlechild, S. C. (1975**)** "Two-part tariffs and consumption externalities", Bell Journal of Economics and Management Science. Vol. 6, No 2, 661-670.

National Telecommunications and Information Administration, NTIA, (1999) Falling Through the Net: Defining the Digital Divide. Washington DC. https://www.ntia.doc.gov/report/1999/falling-through-net-defining-digital-divide

OECD (2001). "Understanding the Digital Divide", *OECD Digital Economy Papers*, No. 49, OECD Publishing, Paris. http://dx.doi.org/10.1787/236405667766.

Pérez-Hernández, J. and R. Sánchez-Mangas (2011) "To have or not to have Internet at home: Implications for online shopping", Information Economics and Policy 23, 213-226.

Von Rabenau, B. and Stahl, K. (1974). "Dynamic Aspects of Public Goods: A further Analysis of the Telephone System", Bell Journal of Economics and Management Science, Vol 5. No. 2, autumn, 651-669.

Răileanu, M. (2018) "New insights from a multilevel approach to the regional digital divide in the European Union", Telecommunications Policy 42, 452-463.

Robles, J.M. and Torres, C. (2012). "Digital Divide and the Information and Communication Society in Spain". Sociologija i Prostor, 194 (3), 291-307.

Rohlfs, J. (1974). "A Theory of Interdependent Demand for a Communications Service". Bell Journal of Economics and Management Science, vol. 5, 1, 16-37.

Sharma, S. K. (2005). "Socio-Economic Impacts and Influences of E-Commerce in a Digital Economy" in Digital Economy: Impacts, Influences and Challenges, IGI Global, 1-20.

Squire, L. (1973). "Some aspects of optimal pricing for Telecommunications", Bell Journal of Economics and Management Science, Vol. 4, No 2, autumn, 515-525.

Taylor, L. D. (1994). Telecommunications Demand in Theory and Practice. Kluwer Academic Publishers.

Valarezo, Á., Pérez-Amaral, T., Garín-Muñoz, T., Herguera, I., and López, R. (2018). "Drivers and Barriers to Cross-Border E-Commerce: Evidence from Spanish Individual Behavior", Telecommunications Policy, Vol. 42, Issue 6, 464-473.

Valarezo, Á., López, R., Pérez-Amaral, T., Garín-Muñoz, T. and Herguera, I. (2019). Adoption of e-commerce by individuals: Evidence from pooled data in Spain. Forthcoming, Book in Honor of Gary Madden.

Varian, H.R. (2002). "The Demand for Bandwidth: Evidence from the INDEX Experiment". In Robert, C. and James, A. (Eds.). *Broadband: Should We Regulate High-Speed Internet Access?* Washington, D.C.: AEI-Brookings Joint Centre for Regulatory Studies.

Vicente, M.R. and A.J. López (2008). “Some empirical evidence on Internet diffusion in the New Members States and Candidate Countries of the European Union”. *Applied Economic Letters*, 15, 1015-1018.

Appendix

Figures 5-8 represent selected estimated coefficients of the equations in section 4.

Figure 1

Odds of e-commerce adoption. Panel (2008-2016)

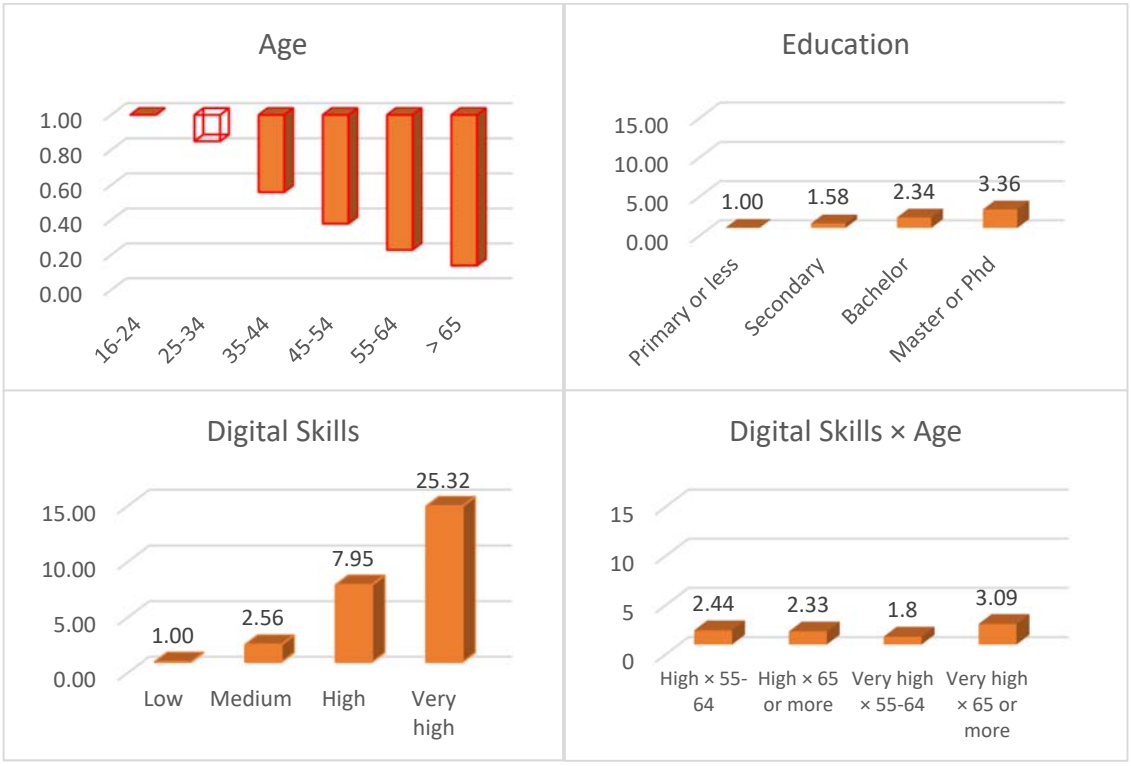


Figure 2:
Odds of e-commerce adoption. Panel (2008-2016)

Figure 3:
Odds ratios of yearly dummies. Panel (2008-2016)

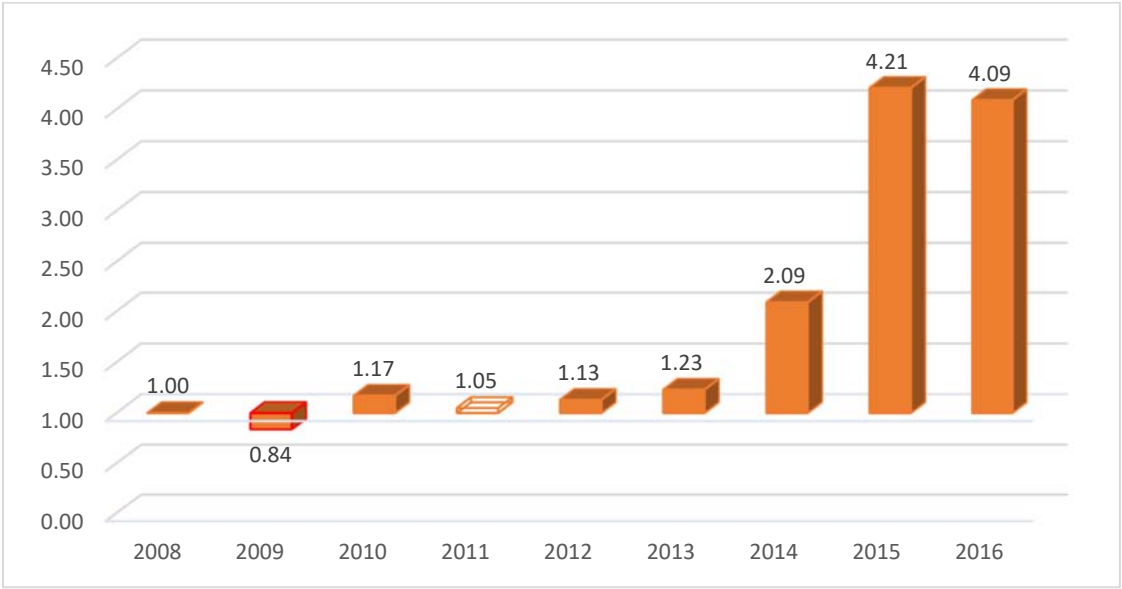


Figure 4:
Odds ratios of e-commerce adoption by regions. Panel (2008-2016)

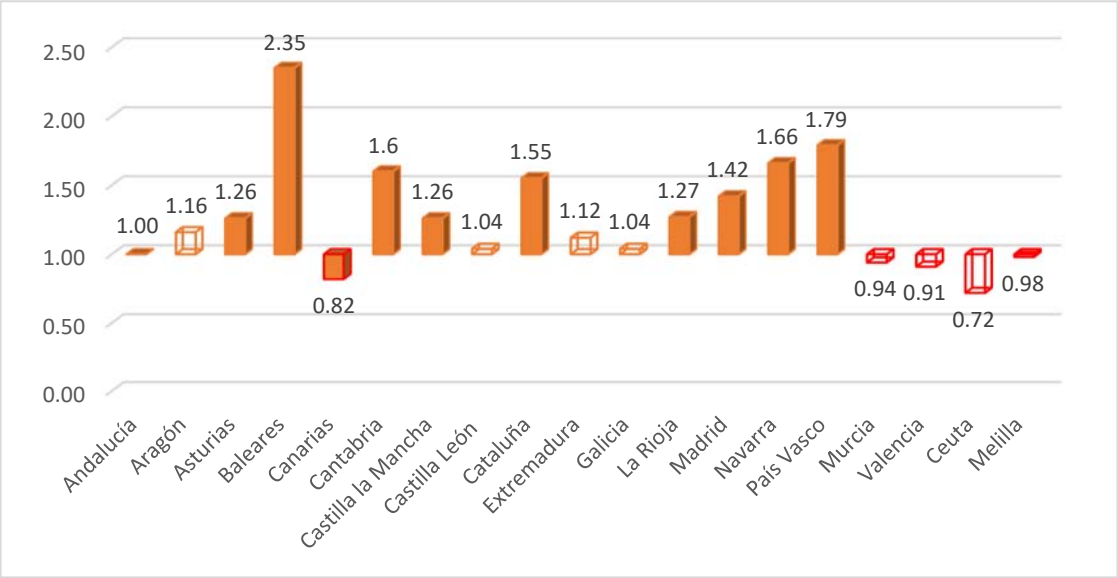


Table A1.
Heckman selection model of adoption of e-commerce by individual Internet users (2008–2016).

	Variables		(3) Internet use. Selection equation. Probit		(4) E-commerce adoption. Probit	
			Coef.	z	Coef.	z
	Gender	Female				
		Male	0.05	1.08	**0.15**	**11.17**
Exclusion restriction	Broadband	No				
		Yes	**1.10**	**21.71**		
Sociodemographic	Age	16-24				
		25-34	0.01	0.08	-0.08	-1.00
		35-44	-0.01	-0.09	**-0.29**	**-3.77**
		45-54	-0.24	-1.69	**-0.43**	**-5.41**
		55-64	**-0.43**	**-2.91**	**-0.63**	**-7.47**
		65<	**-0.40**	**-2.48**	**-0.84**	**-8.67**
	Education	Primary or less				
		Secondary	-0.06	-1.15	**0.19**	**6.70**
		Bachelor	**0.03**	**0.38**	**0.33**	**10.37**
		Master´s/Phd	**0.01**	**0.13**	**0.47**	**14.97**
	Digital Skills	Low				
		Medium	**1.32**	**4.47**	**0.39**	**4.87**
		High	**5.53**	**40.47**	**0.89**	**11.52**
		Very high	**5.45**	**39.65**	**1.44**	**18.67**
	Habitat	<20,000				
		20,000-100,000	**0.12**	**1.98**	-0.03	-1.46
		100,000-500,000	0.03	0.27	**-0.12**	**-4.32**
		500,000<	0.00	-0.02	-0.01	-0.46
	Household Members	One				
		Two	-0.03	-0.51	**-0.07**	**-2.91**
		Three	-0.15	**-2.06**	**-0.15**	**-6.45**
		Four	-0.13	-1.65	**-0.16**	**-6.71**
		Five or more	-0.11	-1.05	**-0.22**	**-6.98**
	Nationality	Foreigner				
		Spanish	-0.10	-1.20	**0.08**	**2.88**
Economic	Employment Situation	Employed				
		Unemployed	**-0.15**	**-2.42**	**-0.14**	**-7.02**
		Retired	**-0.43**	**-5.22**	-0.06	-1.57
		Student	**1.15**	**2.81**	**-0.15**	**-4.64**
		Housekeeper	**-0.24**	**-3.10**	-0.04	-1.34
		Other	**-0.35**	**-3.07**	-0.06	-1.35
	Income	Low				
		Medium	0.08	1.52	**0.21**	**10.41**
		Medium-high	0.12	1.65	**0.43**	**19.67**
		High	0.10	0.99	**0.62**	**23.89**

Table A1. (Continued)

	Variables		(3) Selection equation: Internet use. Probit		(4) E-commerce adoption. Probit	
			Coef.	z	Coef.	z
Interaction	Digital Skills × Age	Medium × 25-34	-0.19	-0.59	0.15	1.64
		Medium × 35-44	-0.16	-0.51	**0.34**	**3.85**
		Medium × 45-54	0.28	0.80	**0.33**	**3.65**
		Medium × 55-64	-0.10	-0.31	**0.34**	**3.63**
		Medium × 65 <	-0.13	-0.40	**0.37**	**3.49**
		High × 25-34	0.06	0.39	**0.24**	**2.70**
		High × 35-44	0.06	0.42	**0.37**	**4.36**
		High × 45-54	0.21	1.39	**0.37**	**4.24**
		High × 55-64	**0.44**	**2.80**	**0.41**	**4.40**
		High × 65 <	**0.46**	**2.70**	**0.40**	**3.65**
		Very high × 25-34	0.09	0.56	**0.27**	**3.02**
		Very high × 35-44	0.01	0.07	**0.43**	**4.97**
		Very high × 45-54	0.19	1.23	**0.42**	**4.72**
		Very high × 55-64	**0.45**	**2.54**	**0.33**	**3.34**
		Very high × 65 <	**0.62**	**2.93**	**0.54**	**4.50**
Time effects	Year	2008				
		2009	-5.90	-1.86	**-0.08**	**-3.67**
		2010	0.08	0.03	**0.07**	**2.79**
		2011	-5.97	-1.86	0.01	0.44
		2012	**-6.07**	**-2.01**	0.03	1.13
		2013	-0.10	-0.03	**0.07**	**2.92**
		2014	-5.87	-1.89	**0.29**	**11.53**
		2015	-0.08	-0.03	**0.60**	**23.96**
		2016	-0.10	-0.03	**0.58**	**23.24**
Sociogeographic effects	Autonomous Communities	Andalucía				
		Aragón	0.15	1.38	0.06	1.56
		Asturias	0.05	0.46	**0.11**	**3.16**
		Baleares	0.27	1.90	**0.35**	**8.78**
		Canarias	0.09	0.74	**-0.09**	**-2.33**
		Cantabria	0.22	1.70	**0.20**	**5.14**
		Castilla la Mancha	0.09	0.81	**0.09**	**2.37**
		Castilla León	**0.36**	**3.08**	0.02	0.52
		Cataluña	0.01	0.10	**0.18**	**6.35**
		Extremadura	0.22	1.92	0.05	1.24
		Galicia	0.17	1.64	0.01	0.41
		La Rioja	**0.27**	**2.02**	**0.10**	**2.42**
		Madrid	0.16	1.56	**0.13**	**4.53**
		Navarra	0.17	1.70	**0.22**	**6.72**
		País Vasco	0.11	0.91	**0.26**	**7.34**
		Murcia	0.01	0.12	-0.02	-0.56
		Valencia	**0.20**	**2.12**	-0.05	-1.51
	Autonomous Cities	Ceuta	0.47	1.39	-0.15	-1.77
		Melilla	0.05	0.15	-0.04	0.15
Constant			6.88	2.23	1.71	-20.26
N. observations			66,099		66,999	
Wald χ^2			4592.85, DF: 70		13666.89, DF: 70	
Pseudo R^2			0.5116			
Wald χ^2: H_0 = independent equations			1.12 DF:1 p-value = 0.2900			

Notes: Coefficients and z statistics significant at the 5% are represented in bold. Probit model, selection equation (3). Probit model, e-commerce adoption equation (4). Heckman probit clustered standard errors. Estimations equation (3) refer to all Internet users that have purchased (or not) online in the last year. The first category of each variable is the reference and is omitted. The samples reflect the fact that some variables are missing for some observations, which reduces the usable sample size. All estimations have been calculated using Stata 15.

Statistical significance and signs of coefficients of all explanatory variables coincide between the probit estimation results (second stage of Heckman selection model), Table A1, and estimations of the two logistic specifications, Table 2. On the other hand, a formal test of independence of two equations of the selection model (equations (3) and (4), with $\chi 2 = 1.12$, and p-value = 0.29) does not provide sufficient evidence to reject the hypothesis of independent equations, implying that, for the considered sample, the initial decision of using the Internet is unrelated to the decision of adopting e-commerce for private use. This may suggest that the modelling approach of using just one step delivers a useful model for e-commerce.