# Online education adoption in Spain 2008-2019. Drivers and impediments.

Rafael López, ICAE, UCM

Ángel Valarezo, ICAE, UCM

Teodosio Pérez-Amaral, ICAE, UCM

- ICAE: Instituto Complutense de Análisis Económico, Spain
  - UCM: Universidad Complutense de Madrid, Spain.

**Acknowledgements**: The collaboration of INE is gratefully acknowledged. Research funded by the Autonomous Community of Madrid, Spain (Project: H2019/HUM-5859, MIMA-CM, Mercados, Innovación y Medio Ambiente), and National R&D Plan, Project: PID2019-107161GB-C32, 2020-2023.

**Keywords**: E-learning adoption, online education, digital skills, linear probability model, logistic model, Heckman model.

**Declaration of interest**: none

**JEL Classification**: C30; L86; L96.

**Abstract**

Online learning and training are gaining momentum worldwide by reducing the temporal and spatial limitations associated with the traditional form of face-to-face education. Online education improves access to education and training, especially during the present Covid-19 pandemic. This article focuses on online education adoption in Spain. A large and representative panel database from the ICT in the household´s survey by the National Institute of Statistics is used. The first objective is to provide an econometric model for the adoption of online education. Next is to measure the effects of relevant observable individual socioeconomic variables on adoption. A Heckman selection model using panel data for 2008-2019 allows estimating the impact of differences in gender, age, education, digital skills, habitat and income. The drivers and impediments have the expected signs and plausible sizes. The paper concludes with policy recommendations and suggestions for further research.

## 1. **Introduction**

Online learning and training are gaining momentum worldwide by reducing the temporal and spatial limitations associated with the traditional form of face-to-face education. Online education improves access to education and training, increases education quality, reduces the cost of learning, and enhances the education's cost-effectiveness (Bates, 1997). Online education is provided mainly in two polar modalities: asynchronous and synchronous environments and various blends of both, including partly online and partly face-to-face activities.

Asynchronous online learning has the advantages of accessing material anytime and anywhere, reaching a greater audience, and achieving uniformity of content. Synchronous learning is closer to the classroom experience while allowing recording for later use. Online learning enables teaching staff can be distributed geographically while obtaining more efficiency by teaming their expertise together. Instructors can successfully combine online activities with face-to-face teaching.

To make precise the terminology, we follow Money & Dean (2019) by considering online education, online delivery, online education design, and computer-based instruction as equivalent. According to Volery & Lord (2000), there are essentially interchangeable terms and phrases that encompass instruction through computer-based, internet-based, and other digital technologies. They also include the words online teaching and learning, e-learning, and distance learning as similar. Simonson et al. (2011) define distance learning as institution-based, formal education in which learners are physically separated but connected with instructors, resources, and other learners by telecommunications systems. Larreamendy-Joerns & Leinhardt (2006) concluded that distance education and online education essentially have merged (see also Walker & Kelly, 2007). The phrase online teaching and learning, as applied by Goodyear et al. (2001), similarly refers to teaching and learning that occur over a computer network and in which interactions among people – including both synchronous and asynchronous forms of interaction – remain an essential part of the learning process.

According to Research and Markets (2018 a, b), the global e-learning market could reach 65.41 billion dollars by 2023.

Despite several advantages of online learning, retaining students is a key challenge, with a high attrition rate (Perna et al., 2014). Online education also presents disadvantages relative to traditional face-to-face learning. The feedback between the instructors and the students is mainly lost, as well as the relationships among the students. The possibility of chatting before, after class, and during breaks is limited. The physical presence of the student in front of the computer in synchronous learning is not assured. The instructor does not know to whom he/she is talking.

Most of the existing literature uses models of technology adoptions for explaining and understanding the decision to adopt online education technology (Panigrahi et al., 2018).

The first objective of the current article is to provide an economic framework for adopting online education. A second objective is to model adoption using a large data set. A third objective is to measure the effects of relevant observable variables on adopting or not online education. This framework may be interesting for online students and instructors. Teaching institutions and national authorities may also benefit from the output of the research.

The rest of the article is as follows. Section 2 contains a literature review. Section 3 shows the theoretical framework. Section 4 includes the presentation and description of the data. The empirical model and estimation results are in section 5. Section 6 contains the conclusions.

## 2. Literature review

The literature on this topic is abundant, with various subtopics, emphasis, and methodologies. Here we concentrate on the adoption of online education. The survey by Panigrahi et al. (2018) analyzes the factors that influence online education's three critical elements: adoption of technology, the continuation of technology use, and e-learning outcomes.

### 2.1 Adoption

Adoption of technology, online consumer behaviour, and the specific decision of online learning adoption has been studied based on several frameworks. The most used approaches are applications, adaptations, and unifications of models of individual acceptance and intention. Among the soundest theoretical backgrounds are Innovation Diffusion Theory (IDT), Rogers, (2003); Theory of Reasoned Action (TRA), Ajzen & Fishbein (1977, 1980); Theory of Planned Behavior (TPB), Ajzen (1991); Technology Acceptance Model (TAM), Davis (1989, 1993); Davis & Warshaw (1989); Theoretical Extension of the Technology Acceptance Model (TAM2), Venkatesh & Davis (2000); Decomposed Theory of Planned Behavior (DTPB) Taylor & Todd, (1995); Expectation-Confirmation Theory (ECT), Oliver (1980); and, the Unified Theory of Acceptance and Use of Technology (UTAUT), Venkatesh et al. (2003).

Most of the rich body of theory and applied research about ICT's diffusion uses IDT, TRA, TAM and TPB, which include variables that affect an individual's motivation to accept new technology and explain the decision-making process of doing so. When the goal is to reach further than the initial acceptance, ECT is widely used. The contributions of this literature are precedents of the current study.

### 2.1.1 Online learning adoption

Panigrahi et al. (2018) examine a bibliographical database, organizing the literature in three categories, technology adoption, the continuation of technology use, and learning outcomes. The antecedents and main extensions of the three categories are discussed extensively. For technology adoption, which is directly related to our research, the discussion focuses on personal and environmental factors. A later survey by Money & Dean (2019) also adopts this strategy.

### 2.1.2. adoption, personal factors

Perceived usefulness (PU) and perceived ease of use (PEoU) are the main factors which, according to the Technology Acceptance Model (Davis, 1989), are the predominant antecedents of technology adoption. Other factors that also affect the acceptance of technology are the perception of interaction, cognitive absorption, self-distraction, cognitive age, social network, national culture, and surrounding conditions (e.g. thermal climate and national wealth). Additional personal factors include perceived behaviour control, performance and effort, expectancies, and user resistance.

### 2.1.3 Adoption, environmental factors

The environmental factors for adoption include perceived characteristics of innovation, subjective norms, facilitating conditions, technology inhibitors, and technology adoption in organizations.

Focusing on French universities, Jacqmin, (2019) analyses the effect of providing MOOCs (Massive Open Online Courses) on new students' enrollment in online and traditional programs and media coverage. The study found that all else being equal, offering MOOCs rises over 2% of the student intakes of universities hosting this kind of programs. A double positive effect is identified, the enhanced information and attention-grabbing.

Bryson & Andres (2020) analyze the effects of the Covid-19 on adopting e-learning in a university. The pandemic forced rapid improvisation and adoption of online teaching. The paper distinguishes between the development of distance learning programs compared to the rapid adoption of online learning. The article concludes that a shift to online education is not about substituting on-campus with online but developing a new and transformational approach that will extend universities' reach and alter their fundamental essence. Part of this shift will reflect the emergence of new bimodal approaches to facilitating learning outcomes accommodating all types of students.

## 2.2 Continuation

While adoption is a necessary component of online learning, the continuation of technology use and e-learning outcomes are also relevant. A summary of these two stages is in the following two sections.

#### 2.2.1 Continuation, personal factors

The personal factors that influence the continuation include satisfaction, habit, flow, and self-efficacy.

#### 2.2.2 Continuation, environmental factors

The main environmental factors that influence the continuation are confirmation, information system quality parameters, psychological safety, communication climate, and perceived responsiveness.

#### 2.2.3 Continuation, knowledge contribution and continuance in virtual communities

The main drivers for continuation related to virtual communities are satisfaction, trust, we-intentions, sense of belongingness, community commitment, immersion, participation needs, attitude, motivations, Knowledge characteristics, interactions, social network integration, structural dynamics, and technology continuance in organizations. Specific research finds that collaborative chat participation in MOOCs slows down the attrition rate over time, Ferschke et al. (2015).

### 2.3 Learning outcome

The learning outcome is the measure of the effectiveness of online education. Success depends on whether it has achieved the desired results. The factors affecting online outcomes are classified as personal and environmental factors as well as contextual differences.

#### 2.3.1 Learning outcome, personal and environmental factors

The personal and environmental factors affecting the learning outcome consist of engagement, motivations, focus, design interventions, virtual competency, team collaboration, contextual differences, formal vs informal learning, virtual world characteristics, cloud computing, and content management system vs learning management system.

Research conducted by Cho et al. (2017) examines the effects of students’ self-regulated learning on their perceptions of community of inquiry and their effective outcomes. Based on the analysis of data of 180 college students, the empirical results show that self-regulated learning levels have meaningful effects on students’ perceived community of inquiry. The authors recommend that instructors and designers use instructional strategies to develop positive learning experiences in online environments.

Money & Dean (2019) contribute with a new content-based literature review and analysis of recently published articles. They report on factors that describe populations of students engaged in formal online programs and courses. The study highlights gaps, inconsistencies in the literature relating to online student populations posing barriers to assessing the impact of population differences on learning outcomes. The authors propose a new conceptual model to guide educators and institutions in delivering more effective online programs. The model identifies critical components to describe student populations, which interact with the main processes that directly or indirectly affect outcomes. They use a methodology that includes a systems-based model by Lowe & Holton (2005), later revised by Knowles et al. (2015). The focus of the study is the outcome rather than the adoption of online education. They do not consider economic factors as possible antecedents for online education.

A recent unpublished study by Sanchis-Guarner et al. (2021) deals with the effect of home high-speed internet on national test scores of students at age 14. They use comprehensive information from relevant sources, as well as econometric techniques. The paper finds that increasing broadband speed by 1Mbit/s increases tests scores by 1.37 percentile ranks in the years 2005-08. This effect is sizeable, equivalent to 5% of a standard deviation in the national score distribution. The result is not driven by other technological mediating factors or school characteristics.

**2.4 Caveats**

The previous literature is extensive and valuable in many ways, but it may help define some additional concepts with precision. There is a need for clarification whether the arguments exposed as antecedents of eLearning adoption are considered from an individual consumer's perspective for private purposes or work purposes. The distinction between the motivations of a personal user and an online education supplier would require further clarification.

The factors considered in the above literature are mostly psychological reasons for adopting online technology. However, the adoption includes a bundle of technology and specific contents, which may be relevant in embracing online education. There is a difference between adopting e-learning technology versus engaging in specific online courses or activities.

The psychological perspective above is very fruitful. However, there is little emphasis on the heterogeneity of the students: their different fields of knowledge, ages, interests, gender, educational background, income levels, geographical locations, and socioeconomic characteristics seem to be somewhat neglected. This characteristic may suggest adopting an economic framework. The decision of adoption is a function of the costs and benefits of online education engagement, based on observable individual and environmental variables. The following section contains an alternative model that accounts for as much heterogeneity as the data can support.

## 3. Theory

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 the consumer surplus associated with Internet usage and the cost of access. Regarding access to online learning, the relevant theory is that of the telecommunications demand framework of Artle & Averous (1973), Squire (1973), von Rabenau & Stahl (1974), Rohlfs (1974), Taylor (1994), Kridel et al. (1999), and Rappoport et al. (2003).

In telecommunications, the use of a specific service is conditional on access to this service (Taylor 1994). The current approach assumes that Internet access is a prerequisite for adopting online education. In any case, a consumer could obtain internet access 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 parlours, 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 2008-2019.

In many cases, access to the internet is not a decision but rather a circumstance governed by the carriers that incorporate the internet without knowledge by the consumer. Demoussis & Giannakopoulos (2006) use a similar argument for the European case using 2002-2003 data.

When access to the internet is widespread, the decision to use the internet no longer needs to be modelled. The hypothesis is that consumers decide to use online education, given that they already have access to the internet[1].

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 & Giannakopoulos (2006); Fairlie (2004); Vicente & López (2008); Lera-Lopez et al. (2011) and (Valarezo et al., 2018, 2020), 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 online education ($Y_i$), will be a function of the benefits $B(Y_i|x)$ of

[1] 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.

doing so and the costs $C(Y_i|x)$, where x is a set of conditioning variables associated with it. The conditional probability of adopting online education 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 individual consumers considered are Internet users. Based on the theoretical model above, an empirical model is specified and estimated in section 5. The following section presents the data set used in this paper.

## 4. The Data

This work uses annual panel data on individuals from the survey on Equipment and Use of Information and Communications Technologies by Households from 2008 to 2019. The survey follows the guidelines of Eurostat and is performed by the Spanish National Statistical Institute (INE, 2020). It is representative at both regional and national levels and includes an elevation factor. The main focus is on the adoption of ICT technology and services by households and individuals. Eurostat coordinates and makes the survey available in all European countries on an annual basis.

Figure 1 shows the penetrations of internet use in the last three months, together with the access to broadband and online education.

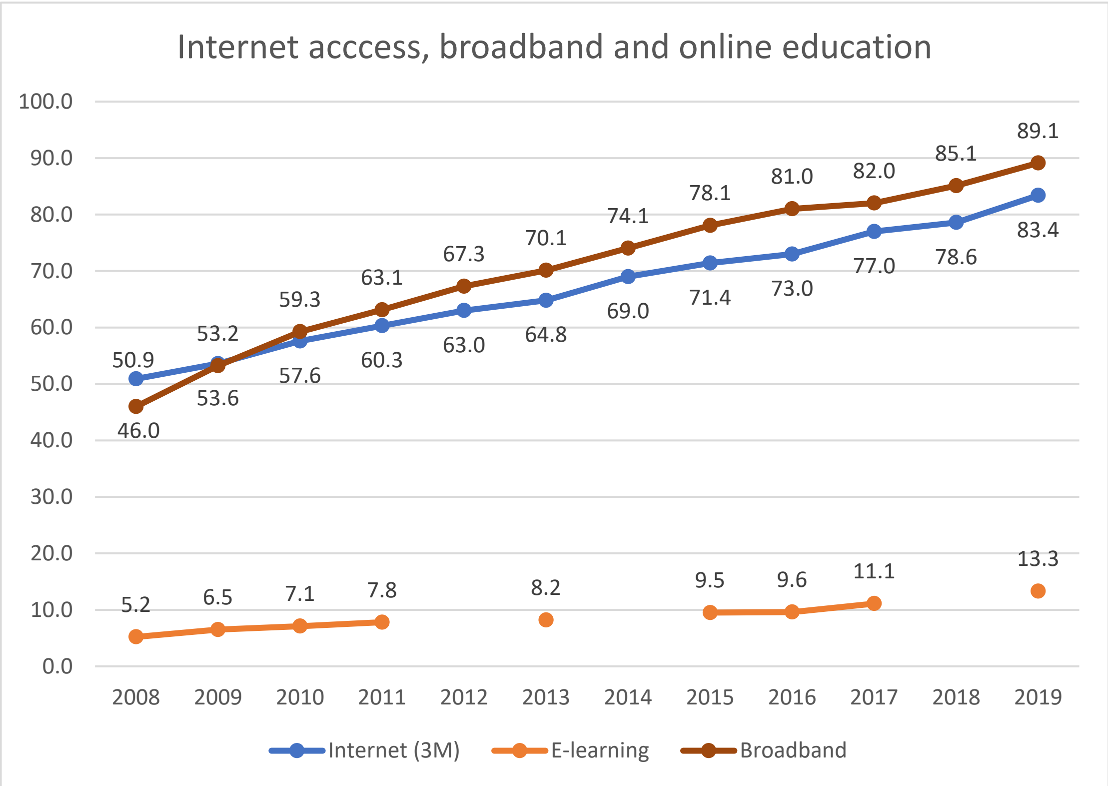


**Figure 1. Access to the internet, broadband, and online education 2008-2019.** Source: Self elaboration based on INE, 2020. Relative to the adult population 16 and above.

Figure 1 shows that Internet access in the last three months and broadband access have increased steadily along the 12 years of the sample, reaching 83.4% and 89.1%, respectively. Online learning has been growing faster, reaching a penetration of 13.3% of the population 16 and above.

In the 12 years of the sample, Internet access has increased by 63.8%, while online learning has increased by 155.7%, which more than doubles the rise of Internet access and shows the rapid increase of online education in Spain.

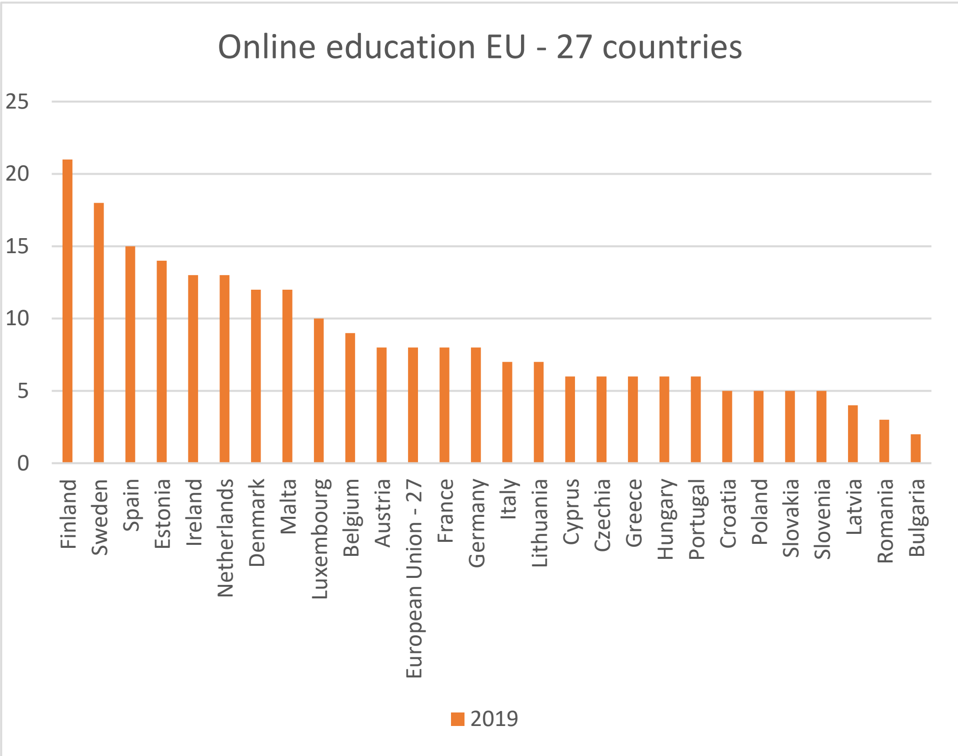


**Figure 2.** Percentage of the population taking an online course. European countries 2019. Source: Self elaboration based on Eurostat, 2020.

Figure 2 shows a comparison with other EU. Spain is in third place in penetration after Finland and Sweden. This position may be due to the existence of a large national university at a distance, UNED (web address), with an enrollment above 250,000 persons in 2019. Other institutions are UOC, UNIr, Isabel de Castilla and several smaller universities and institutions.

### 4.1 Type and Source of microdata.

The primary data are available at INE (2020), containing additional information and reports related to every aspect of the design, implementation, and statistics. It is a rotating survey that includes 18,000-21,000 dwellings each year. The same residence is interviewed for a maximum of four (consecutive) years and a replacement of around 30% every year. Approximately 60% of the interviews are conducted by phone (CATI, computer-assisted telephone interview) and 40% in person (CAPI, computer-assisted personal interview).

The raw panel data underwent rigorous analysis and filtering to extract and homogenize the information about individuals uniquely identified throughout the observation period. The process goes as follows:

### 4.2 Constructing a pool

The information on dwellings for each year is publicly available in INE (2020). Each year includes a raw data file with the socio-demographic information of the members of each residence, a second data file with the responses to the survey, and an additional Excel file with the description of both raw data files.

Next, the variables are subject to a process of homogenization. The survey's questionnaire varies each year according to the evolving situation of ICT. It implies that some new variables emerge, and some old ones disappear. Besides, each year, the names of many variables change (in publicly available microdata, most variables are tagged after their ordinal situation in the questionnaire – which usually changes). To homogenize the variables, we developed a script in R that scrapped the description excel files for identifying, across years, the same variables with different names and different variables that were assigned the same name.

### 4.3 Constructing the panel

The next step converts a pool of dwellings into a panel data set of individuals. This transformation requires incorporating the dwelling identifier supplied by the INE into the previous pool database.

However, we are interested in identifying individuals rather than dwellings. A dwelling that participates in the survey for several years may have different respondents across years and correspond to other family groups. Socio-demographic information for each residence member, like gender and date of birth, was used as a filter. These criteria allow identifying whether the respondent from a specific dwelling was always the same individual or not.

The panel obtained from the survey 2008-2019 is used for the calculations below. It consists of 210,370 observations (corresponding to 97,859 different individuals) and more than 750 variables. Table 1 contains a list and description of the main variables used in this article.

| Dimensions | Characteristics/ Variables | Categories or levels within each variable |
| --- | --- | --- |
| **Services** | **Internet access** | 2 groups: 1 if access; 0 otherwise. |
| | **Internet user** (in the last 3 months) | 2 groups: 1 if internet user; 0 otherwise. |
| | **E-learning** | 2 groups: 1 if e-learning user; 0 otherwise |
| **Sociodemographic** | **Gender** | 2 groups: 1 if male, 0 if female |
| | **Age** | 7 groups: 16-25, 25-35, 35-45, 45-55, 55-65, 65-75, 75 or more. |
| | **Household members** | 5 groups: One, Two, Three, Four, Five or more. |
| | **Population size (Habitat)** | 5 groups: less than 10,000; 10,000-20,000; 20,000-50,000; 50,000-100,000; and 100,000 or more and province capitals. |
| | **Nationality** | 2 groups: 1 if Foreigner; 0 if Spanish. |
| **Individual** | **Education** | 4 levels of study: None or Primary, Secondary, Bachelor's Degree, and Master or PhD. |
| | **Digital Skills** | 4 levels: Low, Medium, High, and Very High. |
| | **Internet trust** | 3 groups: Low, Medium, High. |
| **Economic** | **Income** (monthly net income) | 4 groups: Low, Medium, High, and Very High. |

**Table 1. Dimensions, variables, and categories related to e-learning.**
Source: Self elaboration based on TICH survey of INE (2020).

Table 1 contains, in the first column, the dimensions relevant for the gaps in this study: socio-demographic, individual skills, and economic characteristics. The second column contains the specific variables available in this data set for approximating each dimension. In the third column, the different levels or categories of each variable. The variables are coded as dummies to facilitate interpretation.

Most of the previous variables and dimensions are well-established in the literature. These include Digital Skills, which are relevant according to the European Commission (2019, 2020)[2]. They are crucial elements to measure to promote digital inclusion.

The Digital Skills variable used in this study is a synthetic index based on the former European Commission's Digital Skills Indicator, which accounts for the number and complexity of activities involving the adoption and use of digital equipment and Internet services (European

[2] The other variables considered are also relevant according to European Commission (2019).

Commission, 2016). The index covers four areas of competence: information and data literacy, communication and collaboration, problem-solving, and software skills for content manipulation.

The respondents reveal their capabilities by answering specific questions in each area. Information skills account for people who have copied or moved a file or folder, used internet storage space, searched for information online, searched for information about goods and services, and searched for health-related topics. Communication and collaboration areas are estimated using email, social networks, telephone calls through the internet, and web sites for sharing their content. The problem-solving area of competence regards transferring files between devices, installing software and apps, changing the software settings, selling goods or services through the internet, using online education material, and carrying out online banking activities. The software skills include using text processing software, spreadsheet software, software to edit audio-visual content, creating documents that integrate different files, writing a computer program using specialized programming language, and using a spreadsheet's advanced functions.

This paper uses four categories of digital skills: low, medium, high, and very high. These correspond to the quartiles of the percentage of tasks that an individual consumer declares to perform.

### 4.4 Data description

Table 2 shows the penetration rates of online education according to the different variables from 2008-2019. The data are missing for the years 2012, 2014 and 2018. Internet Trust is not available before 2015.

The penetrations for 2019 reveal no gender gap, the ages with the highest penetration are 25-34 years, and the number of household members is relevant. The population size has an increasing relationship with online education, and Spanish nationals are more inclined to engage in online education than foreigners.

Online education has an increasing and strong relationship with the level of formal education and with Digital Skills. There is also a clear positive relationship between the adoption of online education with internet trust and income.

| Variables | Categories | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gender | Female | 5.3 | 6.5 | 7.4 | 7.7 | | 8.7 | | 9.3 | 9.2 | 10.7 | | 13.2 |
| | Male | 5.1 | 6.4 | 6.9 | 7.9 | | 7.7 | | 9.7 | 9.9 | 11.6 | | 13.3 |
| Age | 16-24 | 7.3 | 8.7 | 9.8 | 10.6 | | 10.4 | | 12.6 | 14.4 | 16.6 | | 16.7 |
| | 25-34 | 10.1 | 11.2 | 12 | 14.4 | | 16.2 | | 16.0 | 16.7 | 19.7 | | 23.7 |
| | 35-44 | 6.3 | 8.9 | 10.4 | 10.1 | | 11.0 | | 14.0 | 13.3 | 15.7 | | 19.1 |
| | 45-54 | 4.6 | 6.5 | 7.0 | 7.7 | | 8.9 | | 9.7 | 10.8 | 13.2 | | 16.2 |
| | 55-64 | 2.0 | 2.5 | 2.8 | 4.0 | | 3.7 | | 6.4 | 5.6 | 6.1 | | 8.4 |
| | 65-74 | 0.4 | 0.8 | 1 | 0.7 | | 0.8 | | 1.9 | 1.3 | 1.5 | | 2.3 |
| | 75+ | 0 | 0.2 | 0 | 0.2 | | 0 | | 0.1 | 0.1 | 0.1 | | 0.4 |
| Household Members | One | 2.7 | 5.0 | 5.0 | 5.7 | | 6.0 | | 6.3 | 7.0 | 7.8 | | 11.3 |
| | Two | 4.0 | 5.3 | 5.1 | 5.8 | | 5.8 | | 7.4 | 7.6 | 8.2 | | 8.8 |
| | Three | 6.0 | 7.4 | 8.1 | 9.0 | | 10.3 | | 10.5 | 9.9 | 13.0 | | 14.9 |
| | Four | 6.8 | 7.3 | 9.3 | 9.7 | | 9.3 | | 11.9 | 12.5 | 13.7 | | 18.7 |
| | Five or more | 4.0 | 6.1 | 5.6 | 7.3 | | 8.7 | | 9.7 | 9.3 | 11.8 | | 12.6 |
| Population Size | 100.000+, Capitals | 5.8 | 6.9 | 8.2 | 9 | | 10.0 | | 11.7 | 10.7 | 12.7 | | 15.3 |
| | 50.000 to 100.000 | 7.1 | 8.0 | 6.8 | 7.1 | | 8.0 | | 9.0 | 11.8 | 11.9 | | 14.6 |
| | 20.000 to 50.000 | 4.6 | 6.8 | 7.4 | 7.9 | | 6.7 | | 7.8 | 9.4 | 10.3 | | 12.7 |
| | 10.000 to 20.000 | 3.8 | 6.0 | 6.3 | 7.0 | | 7.8 | | 9.6 | 8.0 | 10.0 | | 10.5 |
| | Less than 10.000 | 4.3 | 5.1 | 5.6 | 6.4 | | 6.4 | | 6.6 | 6.8 | 8.6 | | 10.2 |
| Nationality | Spanish | 5.2 | 6.8 | 7.2 | 8 | | 8.5 | | 9.7 | 9.8 | 11.4 | | 13.9 |
| | Foreigner | 4.8 | 4.6 | 6.6 | 6.2.0 | | 6.3 | | 7.3 | 7.3 | 8.9 | | 8.0 |
| Education | None or Primary | 0.3 | 0.5 | 0.6 | 0.8 | | 0.5 | | 0.4 | 0.4 | 0.8 | | 1.0 |
| | Secondary | 4.2 | 5.5 | 6.1 | 6.3 | | 6.2 | | 6.8 | 7.0 | 7.6 | | 9.9 |
| | Bachelor's Degree | 7.8 | 10.1 | 10.9 | 11.2 | | 13.5 | | 18.2 | 18.1 | 22.4 | | 22.5 |
| | Master or PhD | 15.7 | 18 | 18.9 | 22 | | 21.2 | | 26 | 25.5 | 27.7 | | 31.0 |
| Digital Skills | Low | 0.3 | 0.2 | 0.2 | 0.4 | | 0.4 | | 0.5 | 0.4 | 0.6 | | 0.6 |
| | Medium | 7.4 | 5.8 | 7.1 | 5.7 | | 5.1 | | 4.9 | 5.1 | 5.7 | | 6.0 |
| | High | 23.4 | 18.8 | 22.6 | 17 | | 17.4 | | 17.3 | 15.2 | 16.7 | | 17.9 |
| | Very High | 55 | 42.7 | 48.2 | 42.1 | | 44.5 | | 44.3 | 45.6 | 48.2 | | 49.9 |
| Internet Trust | Low | | | | | | | | 8.9 | 10.5 | 9.0 | | 9.4 |
| | Medium | | | | | | | | 14.3 | 13.9 | 16.5 | | 17.8 |
| | High | | | | | | | | 20.6 | 16.2 | 20.8 | | 28.3 |
| Income | Low | 1.6 | 1.8 | 2.6 | 3.4 | | 2.8 | | 4.6 | 3.9 | 4.7 | | 6.1 |
| | Medium | 4.1 | 5.1 | 6.4 | 6.6 | | 7.0 | | 6.2 | 6.3 | 7.7 | | 10.6 |
| | High | 7.4 | 10.2 | 10.4 | 12.4 | | 12.7 | | 15 | 14.3 | 16.0 | | 18.3 |
| | Very High | 12.5 | 14.6 | 17 | 17.5 | | 19.6 | | 23.7 | 21.4 | 25.1 | | 26.6 |

**Table 2. Online education penetration rates by variables 2008-2019. Blanks mean not available.**
Source: Self elaboration based on TICH survey of INE (2020).

Figure 3 shows the gaps in the penetrations of online education according to the main variables. The gaps are computed as the difference between the category with the highest penetration and the other categories divided by the penetration of the highest category (Pérez-Amaral et al., 2021), paper on divides.

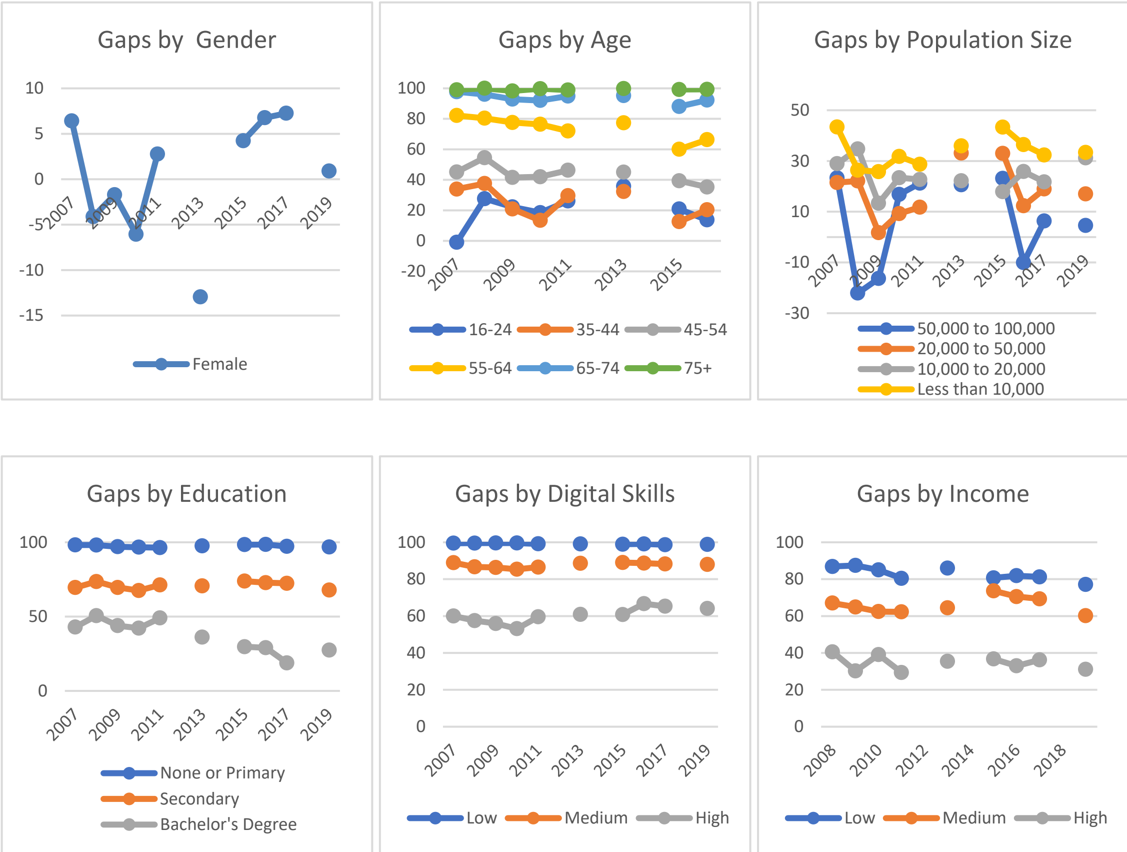


**Figure 3. Gaps of adopting online education by socioeconomic variables**
Source: Authors' elaboration from INE (2020).
Notes: Data were not available for online education (2012, 2014 and 2018). Reference categories: gender: male; age, 25-34; population size, 100,000+ and provincial capitals; education: master or PhD; digital skills: high digital skills; income: very high income.

Figure 3 shows that the gender gap is almost zero in 2019, while the gaps by age are significant and do not decrease over time.

The gaps by population size are present with a 30% gap for sizes of less than 10,000 inhabitants compared to populations of more than 100,000 inhabitants and provincial capitals.

The gaps in online learning by education are large and do not decrease over time. In particular, the gap for none or primary studies is close to 100% compared to masters or PhD degrees, and the gap for the group with secondary studies is 68%.

The gaps related to digital skills are also large and stable. They go from close to 100% for low digital skills to around 60% for high digital skills (compared to very high digital skills).

The gaps by income are also large and stable, ranging from 80% for low income to 30% for high income (compared with very high income).

According to their main explanatory variables, many of the gaps are large and do not show a clear decreasing pattern. This persistence leads us to the need of modelling these relationships in a synthetic and interpretable manner.

Table 2 and Figure 3 indicate which variables may be relevant for adopting online learning in Spain. Still, the relationships are only indicative since the effect of each variable is mixed with those of other possible explanatory variables.

A high correlation between explanatory variables may be problematic in terms of the efficiency of the estimators and hypothesis testing. Table 3 contains the polychoric (Drasgow, 1988; Kolenikov, 2016) correlation matrix between the possible explanatory variables. Polychoric correlations measure the pairwise correlations between a set of variables allowing for discrete dummy variables. The general conclusion is that there are low correlations between the variables considered.

| | Elearn. | Int. Access | Int. User | Gender | Age | Hh. Memb. | Pop. Size | Nation | Educ | Dig. Skills | Int. Trust |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Elearn. | 1 | | | | | | | | | | |
| Int. Access | 0.30 | 1 | | | | | | | | | |
| Int. User | 0.90 | 0.48 | 1 | | | | | | | | |
| Gender | -0.02 | 0.00 | -0.02 | 1 | | | | | | | |
| Age | -0.19 | 0.05 | -0.18 | 0.02 | 1 | | | | | | |
| Hh.Memb. | 0.03 | 0.14 | 0.07 | 0.02 | -0.31 | 1 | | | | | |
| Pop. Size | -0.08 | -0.12 | -0.10 | 0.03 | -0.05 | 0.07 | 1 | | | | |
| Nation | -0.15 | -0.24 | -0.03 | -0.03 | -0.16 | 0.06 | -0.01 | 1 | | | |
| Educ | 0.40 | 0.25 | 0.33 | -0.05 | -0.06 | -0.01 | -0.16 | -0.14 | 1 | | |
| Dig. Skills | 0.65 | 0.40 | 0.77 | 0.08 | -0.37 | 0.08 | -0.12 | -0.16 | 0.52 | 1 | |
| Int. Trust | 0.19 | 0.18 | 0.30 | 0.07 | -0.14 | 0.05 | -0.05 | -0.07 | 0.22 | 0.36 | 1 |
| Income | 0.24 | 0.41 | 0.29 | 0.07 | 0.06 | 0.15 | -0.12 | -0.29 | 0.43 | 0.37 | 0.19 |

**Table 3: Polychoric correlation matrix**
Source: Self elaboration based on TICH survey of INE (2020).

The following section contains a multivariate model that incorporates the possibly explanatory variables discussed so far. In that way, we intend to disentangle the effect that each explanatory variable may have. We look for the partial effect of each of the explanatory variables.

## 5. The Model and Estimation

This section presents an estimable model of the decision of adopting online education. First, for being an online learner, the individual must be an internet user (those who answered using the internet at least once in the last three months).

Figure 4 illustrates the decisions made by individuals when confronted with the choices of using the internet and of using e-learning.

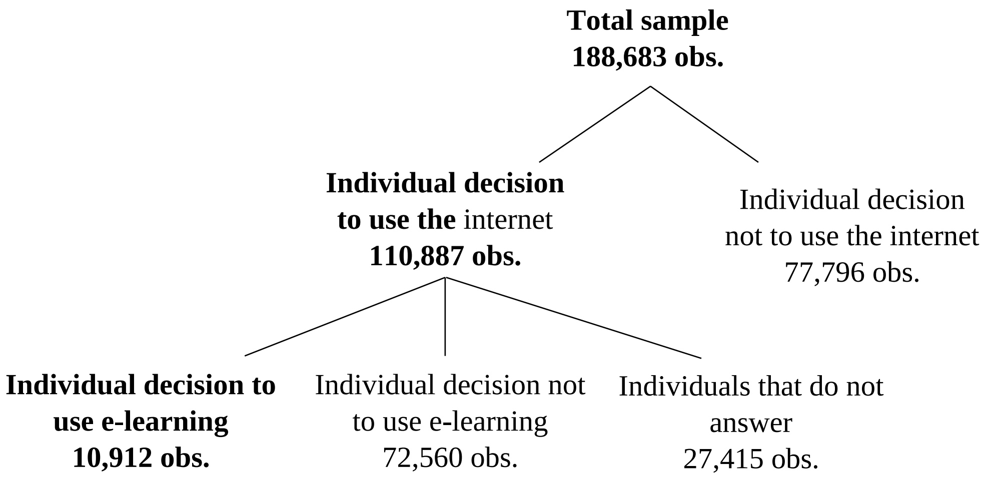


**Figure 4. The decision process of the internet and online education adoption 2008–2019.**

The decision of adopting online education for personal reasons is represented by the binary dependent variable defined below:

**Online learning** = 1, if the individual engaged in online learning in the last 12 months.

= 0, otherwise.

Explanatory variables are grouped as follows: Sociodemographic, Individual Skills, and Economic as shown in Table 1:

The interpretation of the variables goes as follows:

- Gender allows for a gap in the adoption of online learning between males and females,
- Age affects the costs and benefits of adoption, depending on the age ranges,
- Habitat increases or reduces the benefits of adopting online education, depending on the size of the populations to which the individual belongs,
- Household Members may affect the benefits and the costs of online education.
- Nationality might signal effects dependent on costs (e.g., different language) and possible benefits (e.g., access to a broader range of services, education in your language).

- Education and Digital Skills[3] are expected to diminish the costs of using online education as well as signal higher potential benefits.
- Economic variables: Income is supposed to increase the benefits of online education.

The following is the basis for an estimable model of the decision of adopting online education:

$$P(Online\_education_{it}) = f\,(Gender_{it}, Age_{it}, Education_{it}, Digital\ Skills_{it}, Income_{it}) + u_{it} \qquad (1)$$

The subindex i stands for individuals, while t stands for time. $P(Online\_education_{it})$ is the probability of an individual i adopting online education at time t and f es a linear (linear probability model, LPM) or logistic function, depending on the model adopted.

The estimation results are presented in table 4 and figure 5. The first column of table 4 contains the names of the variables. The second includes the categories of each variable. The following columns marked with (1) show the estimates and z statistics of the linear probability model of adoption. The columns marked with (2) contain the results for the logit model of adoption, and the last two columns marked with (3) include the estimates and z statistics for the adoption equation of the Heckman model. The selection equation of the Heckman model is deferred to the Appendix to facilitate the comparison between the three adoption equations (1-3).

[3] Digital Skills is a self-elaborated index, based on the answers where the respondent declares whether he or she used specific internet services and/or performed specific computer and internet related tasks.

**Table 4**
**Models of adoption of online education by individual internet users. Panel data (2008-2019)**

| Dep. Variable: | **Online education** | (1) Linear probability model | | (2) Logistic | | (3) Heckman | |
|---|---|---|---|---|---|---|---|
| | | Coef. | z | Coef. | z | Coef. | z |
| Gender | Male | **-0.033** | **-13.30** | **-0.442** | **-12.22** | **-0.030** | **-11.96** |
| Age | 25-34 | **0.022** | **4.380** | **0.225** | **3.280** | 0.007 | 1.010 |
| | 35-44 | **0.016** | **3.320** | **0.150** | **2.280** | -0.005 | -0.710 |
| | 45-54 | **0.026** | **5.280** | **0.303** | **4.440** | 0.000 | -0.040 |
| | 55-64 | 0.003 | 0.620 | -0.004 | -0.060 | **-0.031** | **-4.240** |
| | 65-74 | **-0.019** | **-2.980** | **-0.622** | **-5.300** | **-0.062** | **-7.450** |
| | 75+ | -0.015 | -1.590 | **-0.853** | **-3.460** | **-0.143** | **-10.43** |
| Education | Secondary | -0.004 | -0.940 | **0.227** | **2.240** | **0.042** | **10.260** |
| | Bachelor's Degree | **0.030** | **5.670** | **0.708** | **6.690** | **0.092** | **17.180** |
| | Master or PhD | **0.066** | **12.350** | **1.039** | **9.830** | **0.127** | **23.120** |
| Digital Skills | Medium | **0.039** | **11.760** | **1.754** | **20.430** | **0.184** | **33.230** |
| | High | **0.154** | **42.400** | **3.183** | **35.820** | **0.289** | **47.840** |
| | Very High | **0.439** | **90.280** | **5.097** | **49.560** | **0.565** | **71.090** |
| Income | Medium | 0.000 | 0.130 | 0.008 | 0.140 | **0.017** | **5.170** |
| | High | 0.000 | -0.100 | 0.006 | 0.110 | **0.023** | **6.110** |
| | Very High | 0.009 | 1.910 | 0.098 | 1.530 | **0.032** | **6.520** |
| | Constant | 0.013 | 1.970 | -5.832 | -40.05 | -0.195 | -21.09 |
| Observations | | 70,616 | | 70,616 | | 138,879 | |
| Selected | | | | | | 70,616 | |
| Non-selected | | | | | | 68,263 | |
| Groups | | 45,742 | | 45,742 | | 78,224 | |
| Wald $\chi^2$ | | 14,074.34; DF: 16 | | 4,136.24; DF: 16 | | 9,067.62; DF: 16 | |

Notes: Coefficients and z statistics (significant at 5%) are represented in bold. Equation 1 is a random-effects linear probability model, equation 2 is a logistic model, and equation 3 incorporates the adoption equation of Heckman's selection model. The selection equation is reported in the Appendix. Base categories are female, age 16-24, primary or no education, low digital skills, and low income.

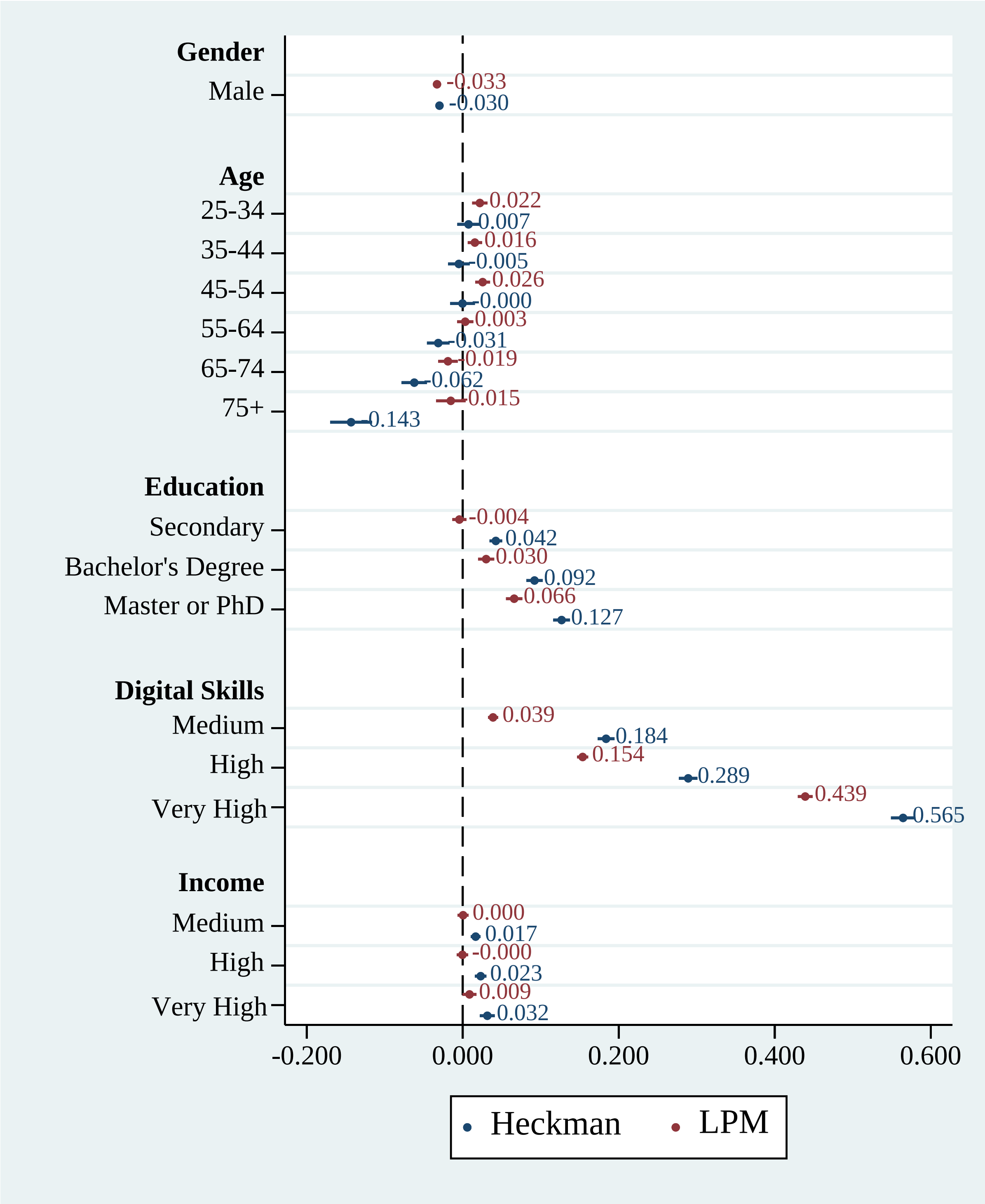


**Figure 5. Models of adoption of online learning by individual internet users. Coefficients and confidence intervals of the linear probability and Heckman models. Panel 2008-2019.**

Notes: Coefficients are depicted by circles and 95% confidence intervals by attached lines. Base categories are female, age 16-24, primary or non-education, low digital skills, and low income. LPM in red. Heckman in blue.

### 5.1 Linear Probability Model (1)

The linear probability model (LPM) constitutes a convenient approach to modelling a binary variable and its determinants. It may suffer the limitation that a predicted probability could lie outside the [0, 1] interval as well as heteroskedasticity. First, these problems can be solved by truncating to 0 or 1 predictions below 0 or above 1 (Rothman, 1986). Heteroskedasticity can be treated using a heteroskedasticity consistent variance matrix (White, 1980).

However, the linear probability model shows reliable estimates of signs and the significance of the coefficients. Moreover, it provides a straightforward interpretation of the estimated coefficients in terms of differences in predicted probabilities (Aldrich & Nelson, 1984; Horrace & Oaxaca, 2006; Hellevik, 2009; Rothman, 1986).

Each significant coefficient of the independent variables in the LPM represents changes in the probability that a person adopts online education for private use and everything else held constant.

Being male decreases the probability of adopting e-learning by 3.3%. Belonging to age groups 25-34, 35-44 and 45-54 increases the probability by 2.2%, 1.6%, and 2.6%, respectively, while being in the 65-74 age group reduces the probability by 1.9%, everything else held constant.

Higher levels of education are related with higher probabilities of adoption, going up by 3% for Bachelor and 6.6% for Master or PhD. Higher levels of digital skills increase the probability of adopting e-learning by 3.9%, 15.4%, and 43.9% for medium, high, and very high. The different levels of income do not appear significant.

### 5.2 Logistic Model (2)

The logistic model is an alternative to the LPM. The logistic produces predicted probabilities in the interval [0, 1] by employing a nonlinear functional form widely used in the literature (Maddala, 1986; Manski & McFadden, 1981).

The coefficients and z statistics confirm the previous findings of the LPM. They are not directly comparable. However, the signs and significance are congruent across the two models.

Being male is associated with a lower probability of adopting online education. Belonging to age groups 25-34, 35-44 and 45-54 increases the probability, while being in the 65-74 and 75+ age group reduces the probability, everything else held constant.

The higher the level of education, the higher the likelihood of adoption. Higher levels of digital skills increase the probability of using e-learning. Differences in income also appear insignificant in this specification.

### 5.3 Heckman selection model (3)

The Heckman selection model (Heckman, 1977), shown in table 3, column (3), uses a two-stage approach to the modelling. First, it uses a selection equation for the decision of being an internet user, shown in table 1A of the Appendix[4]. Second, it uses another equation to explain the decision to the adoption of online education.

The second equation of the adoption is comparable to equations (1) and (2). Male decreases the probability by 3%. Belonging to age groups of 55-64, 65-74, and more than 75 reduces the probability of adoption in 3.1%, 6.2% and 14.3% respectively; and for the groups 25-34, 35-44, and 45-54, there is no evidence in this model and sample that the differences are significantly distinct from zero.

Higher education categories are positively related to online education, increasing the probability of adoption by 4.2% for secondary education, 9.2% for bachelor's degree, and 19.7 % for master and PhD.

Higher levels of digital skills increase the probability of adopting e-learning by 18.4%, 28.9%, and 56.9% for medium, high, and very high levels. Income levels are significant but small and increase the probability of using e-learning by 1.7%, 2.3%, and 3.2%, respectively.

**5.5 The linear probability model and the Heckman model**

Figure 4 compares the point estimates and the confidence intervals for the LPM and the Heckman models. It is apparent that in many cases the Heckman's estimates are higher in absolute values. In particular, in education, digital skills, and income, the effects are more substantial than in the linear probability model.

For gender, the effect is similar in both cases, around -3%. The results are insignificant only in the Heckman model for the groups of age between 25 and 54. In contrast, for higher ages, they are large, negative, and significant in both models.

The estimated values for education are higher, positive, and significant for the Heckman model than the LPM. Digital skills also have a large and significant effect on the adoption of online education, and again Heckman model has larger effects than the LPM. In the case of income, the Heckman model can detect positive and significant effects while the LPM could not.

The previous discussion highlights that models 1-3 give results that are similar in spirit, although they are not the same. The Heckman model is preferred because it incorporates more structure and information and can measure the effects of potentially relevant variables such as income more

[4] The selection equation models the probability of an individual to adopt the use of the internet at a given point in time. This probability depends on the individual characteristics shown in equation A1 of the Appendix.

accurately. Moreover, the three models illustrate that the estimation results are not peculiar to the specification adopted.

## 6. Conclusions and recommendations

This section contains four subsections: general conclusions, policy recommendations, caveats, and further research.

### 6.1 General conclusions.

This article focuses on identifying and measuring the effects of the drivers and impediments on adopting online learning in Spain in the period 2008-2019. It employs the ICT-H survey of the National Institute of Statistics, which allows constructing a panel database of 210,370 individuals.

Using a well-established economic model for online education adoption, we estimate linear probability, logistic, and Heckman selection models. The Heckman model is the preferred specification. The estimations reveal the importance of gender, age, education, digital skills, and income in adopting online education.

The results reveal that being male (-3.0%) and older (up to -14.3%) are impediments to adopting online education (relative to the basis category). Whereas being more educated (up to +12.7%), having better digital skills (up to +56.5%), and higher-income (up to +3.2%) are drivers for adopting online education.

A general conclusion is that the adoption of online education depends on individual factors that can help explain and understand the behaviour related to the adoption of online education. Out of all the variables, the one that makes the most difference is Digital Skills.

### 6.2 Policy Recommendations

Digital Skills causes a sizeable divide in the adoption of online education, apparent in tables 2 and 4. It is the variable that causes the most impact on the probability of adoption. It is the only one that can be influenced by the policy in the short or medium run. The ICT policy would have to focus on this variable.

The European Commission often points to this variable as the main reason why there are low levels of adoption and divides into many services in Spain ((European Commission, 2019). A variety of initiatives about the improvement of Digital Skills are in place at this moment, at the initiative of the European Commission, the Central government, and Autonomous governments (Garín-Muñoz et al., 2019; Valarezo et al., 2020).

Other initiatives may be related to a specific digital divide, like disfavored children, addressed in Garín-Muñoz et al. (2020) and Red.es (2021). They propose to supply computers, tablets, and internet connections to disfavored children to enable them to participate in online education, which has been mandatory in many territories for several weeks or months at a time during 2020 and 2021. This initiative is being implemented currently in several autonomous communities by Red.es (2021) and is expected to cover all geographical areas.

### 6.3 Caveats

The database for this paper is extensive and representative of the whole population of Spain and autonomous communities; however, it is declarative. This problem may introduce recollection bias in the estimates.

The data may not adequately capture the surge in online education in March, April and May of 2020. The survey refers to the respondent person (adult). Meanwhile, children and higher education students were obliged to adopt online learning overnight. Data on the continuation and outcome of online education is not available in this survey.

### 6.4 Further Research

The assessment of the effect of covid-19 on online education in Spain is a promising area of research. The 2020 data for this ICT-H survey are partially helpful since they were collected during April, May, and June of 2020. This database covers until right after the first wave of the pandemic but before the second and third. An evaluation of the total effects of the pandemic will need to wait for the data for the 2021 survey (collected up to June 2021) are available, which will happen on October 25th, 2021.

A comparison with other models that use different paradigms may be helpful. This would include IDT, Rogers, (2003), Theory of Reasoned Action (TRA), Ajzen & Fishbein (1977, 1980), Theory of Planned Behavior (TPB), (Ajzen, 1991), and Technology Acceptance Model, TAM (Davis (1989, 1993); Davis & Warshaw (1989)). Researchers can compare the models based on their policy recommendations.

## References


Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*(5), 888–918. https://doi.org/10.1037/0033-2909.84.5.888

Ajzen, I., & Fishbein, M. (1980). Understanding attitude and predicting social behavior. NJ: Prentice-Hall. *Prentice-Hall*.

Aldrich, J. H., & Nelson, F. D. (1984). *Linear Probability, Logit, and Probit Models*. SAGE.

Artle, R., & Averous, C. (1973). The Telephone System as a Public Good: Static and Dynamic Aspects. *The Bell Journal of Economics and Management Science*, *4*(1), 89–100. JSTOR. https://doi.org/10.2307/3003140

Bates, A. W. (Tony). (1997). The impact of technological change on open and distance learning. *Distance Education*, *18*(1), 93–109. https://doi.org/10.1080/0158791970180108

Bryson, J. R., & Andres, L. (2020). Covid-19 and rapid adoption and improvisation of online teaching: Curating resources for extensive versus intensive online learning experiences. *Journal of Geography in Higher Education*, *44*(4), 608–623. https://doi.org/10.1080/03098265.2020.1807478

Cho, M.-H., Kim, Y., & Choi, D. (2017). The effect of self-regulated learning on college students' perceptions of community of inquiry and affective outcomes in online learning. *The Internet and Higher Education*, *34*, 10–17. https://doi.org/10.1016/j.iheduc.2017.04.001

Davis, F. D., & Warshaw, P. R. (1989). User acceptance of information technology: System characteristics, user perceptions, and behavioral impacts. *Management Sciences*, *35:8*, 982–1003.

Davis, Fred D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Demoussis, M., & Giannakopoulos, N. (2006). Facets of the digital divide in Europe: Determination and extent of internet use. *Economics of Innovation and New Technology*, *15*(3), 235–246. https://doi.org/10.1080/10438590500216016

Drasgow, F. (1988). Polychoric and Polyserial Correlations. In *Encyclopaedia of Statistical Sciences*. John Wiley & Sons, Inc. https://doi.org/10.1002/0471667196.ess2014.pub2

European Commission. (2016). *A new comprehensive Digital Skills Indicator*. Digital Single Market. https://ec.europa.eu/digital-single-market/en/news/new-comprehensive-digital-skills-indicator

European Commission. (2019). *Digital Economy and Society Index (DESI). 2019 Country Report. Spain* [Text]. https://ec.europa.eu/digital-single-market/en/scoreboard/spain

Fairlie, R. W. (2004). Race and the Digital Divide. *Contributions in Economic Analysis & Policy*, *3*(1). https://doi.org/10.2202/1538-0645.1263

Ferschke, O., Yang, D., Tomar, G., & Rosé, C. P. (2015). Positive Impact of Collaborative Chat Participation in an edX MOOC. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), *Artificial Intelligence in Education* (pp. 115–124). Springer International Publishing. https://doi.org/10.1007/978-3-319-19773-9_12

Garín-Muñoz, T., López, R., Pérez-Amaral, T., Herguera, I., & Valarezo, A. (2019). Models for individual adoption of eCommerce, eBanking and eGovernment in Spain. *Telecommunications Policy*, *43*(1), 100–111. https://doi.org/10.1016/j.telpol.2018.01.002

Garín-Muñoz, T., Pérez-Amaral, T., & Garín-Muñoz, M. (2020, May 11). Brecha digital infantil y el COVID-19. *Nada es Gratis*. https://nadaesgratis.es/admin/brecha-digital-infantil-y-el-covid-19

Goodyear, P., Salmon, G., Spector, J. M., Steeples, C., & Tickner, S. (2001). Competences for Online Teaching: A Special Report. *Educational Technology Research and Development*, *49*(1), 65–72.

Hellevik, O. (2009) Linear versus logistic regression when the dependent variable is a dichotomy, Qual Quant 43:59–74 DOI 10.1007/s11135-007-9077-3

Heckman, J. J. (1977). *Sample Selection Bias As a Specification Error (with an Application to the Estimation of Labor Supply Functions)* (Working Paper No. 172). National Bureau of Economic Research. https://doi.org/10.3386/w0172

Horrace, W. C., & Oaxaca, R. L. (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters*, *90*(3), 321–327. https://doi.org/10.1016/j.econlet.2005.08.024

INE, I. N. de E. (2020). *Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares / Últimos datos*. INE. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=estadistica_C&cid=1254736176741&menu=ultiDatos&idp=1254735976608

Jacqmin, J. (2019). Providing MOOCs: A FUN way to enroll students? *Information Economics and Policy*, *48*, 32–39. https://doi.org/10.1016/j.infoecopol.2018.10.002

Knowles, M. S., III, E. F. H., Swanson, R. A., & Robinson, P. A. (2015). *The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development*. Routledge.

Kolenikov, S. (2016). *Polychoric, by any other 'namelist.'* https://www.stata.com/meeting/chicago16/slides/chicago16_kolenikov.pdf

Kridel, D. J., Rappoport, P. N., & Taylor, L. D. (1999). An Econometric Study of the Demand for Access to the Internet. In D. G. Loomis & L. D. Taylor (Eds.), *The Future of the Telecommunications Industry: Forecasting and Demand Analysis* (pp. 21–41). Springer US. https://doi.org/10.1007/978-1-4615-4643-6_3

Larreamendy-Joerns, J., & Leinhardt, G. (2006). Going the Distance With Online Education. *Review of Educational Research*, *76*(4), 567–605. https://doi.org/10.3102/00346543076004567

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

Lowe, J. S., & Holton, E. F. (2005). A Theory of Effective Computer-Based Instruction for Adults. *Human Resource Development Review*, *4*(2), 159–188. https://doi.org/10.1177/1534484305276301

Maddala, G. S. (1986). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

Manski, C. F., & McFadden, D. (1981). *Structural analysis of discrete data with econometric applications*. Cambridge, MA: MIT press.

Money, W. H., & Dean, B. P. (2019). Incorporating student population differences for effective online education: A content-based review and integrative model. *Computers & Education*, *138*, 57–82. https://doi.org/10.1016/j.compedu.2019.03.013

Oliver, R. L. (1980). A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. *Journal of Marketing Research*, *17*(4), 460–469. https://doi.org/10.2307/3150499

Panigrahi, R., Srivastava, P. R., & Sharma, D. (2018). Online learning: Adoption, continuance, and learning outcome—A review of literature. *International Journal of Information Management*, *43*, 1–14. https://doi.org/10.1016/j.ijinfomgt.2018.05.005

Pérez-Amaral, T., Valarezo, A., López, R., & Garín-Muñoz, T. (2021). Digital divides across consumers of internet services in Spain using panel data 2007–2019. Narrowing or not? *Telecommunications Policy*, *45*(2), 102093. https://doi.org/10.1016/j.telpol.2020.102093

Perna, L. W., Ruby, A., Boruch, R. F., Wang, N., Scull, J., Ahmad, S., & Evans, C. (2014). Moving Through MOOCs: Understanding the Progression of Users in Massive Open Online Courses. *Educational Researcher*, *43*(9), 421–432. Scopus. https://doi.org/10.3102/0013189X14562423

Rappoport, P., Kridel, D. J., Taylor, L. D., & Duffy-Demo, K. T. (2003). Residential Demand for Access to the Internet. *The International Handbook of Telecommunications Economics*, *1*, 55–72.

Red.es. (2021). *Programa "Educa en Digital."* https://www.red.es/redes/gl/node/16592

Research and Markets. (2018). *Learning Management System (LMS) Market to 2025 Global Analysis and Forecasts*. https://www.researchandmarkets.com/reports/4436812/learning-management-system-lms-market-to-2025

Rogers, E. M. (2003). *Diffusion of Innovations, 5th Edition*. Simon and Schuster.

Rohlfs, J. (1974). A Theory of Interdependent Demand for a Communications Service. *The Bell Journal of Economics and Management Science*, *5*(1), 16–37. JSTOR. https://doi.org/10.2307/3003090

Rothman, K.J. (1986) Modern Epidemiology. Little, Brown and Company, Boston.

Sanchis-Guarner, R., Montalbán, J., & Weinhardt, F. (2021). *Home Broadband and Human Capital Formation* (SSRN Scholarly Paper ID 3772087). Social Science Research Network. https://papers.ssrn.com/abstract=3772087

Simonson, M., Schlosser, C., & Orellana, A. (2011). Distance education research: A review of the literature. *Journal of Computing in Higher Education*, *23*(2), 124. https://doi.org/10.1007/s12528-011-9045-8

Squire, L. (1973). Some Aspects of Optimal Pricing for Telecommunications. *The Bell Journal of Economics and Management Science*, *4*(2), 515–525. JSTOR. https://doi.org/10.2307/3003051

Taylor, L. D. (1994). *Telecommunications Demand in Theory and Practice*. Springer Netherlands. https://www.springer.com/us/book/9780792323891

Taylor, S., & Todd, P. A. (1995). Understanding Information Technology Usage: A Test of Competing Models. *Information Systems Research*, *6*(2), 144–176. https://doi.org/10.1287/isre.6.2.144

Valarezo, Á., López, R., & Pérez Amaral, T. (2020). Adoption of E-Commerce by Individuals and Digital-Divide. In J. Alleman, P. N. Rappoport, & M. Hamoudia (Eds.), *Applied*

*Economics in the Digital Era: Essays in Honor of Gary Madden* (pp. 103–134). Springer International Publishing. https://doi.org/10.1007/978-3-030-40601-1_4

Valarezo, Á., Pérez-Amaral, T., Garín-Muñoz, T., Herguera García, I., & López, R. (2018). Drivers and barriers to cross-border e-commerce: Evidence from Spanish individual behavior. *Telecommunications Policy*, *42*(6), 464–473. https://doi.org/10.1016/j.telpol.2018.03.006

Varian, H. R. (2002). The Demand for Bandwidth: Evidence from the INDEX Project. *Crandall, R.W., and Alleman, J.H. (Eds.) Broadband: Should We Regulate High-Speed Internet Access? Aei-Brookings Joint Center for Regulatory Studies, American Enterprise Institute*, 39–56.

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Vicente, M. R., & López, A. J. (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.

Volery, T., & Lord, D. (2000). Critical success factors in online education. *International Journal of Educational Management*, *14*(5), 216–223. https://doi.org/10.1108/09513540010344731

von Rabenau, B., & Stahl, K. (1974). Dynamic Aspects of Public Goods: A Further Analysis of the Telephone System. *The Bell Journal of Economics and Management Science*, *5*(2), 651–669. JSTOR. https://doi.org/10.2307/3003127

Walker, E. C., & Kelly, E. (2007). Online instruction: Student satisfaction, kudos, and pet peeves. *Quarterly Review of Distance Education*, *8*(4), 309–319. Scopus.

White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica. 48-4, 817-838.

# Appendix

**Table A1**
**Internet user selection equation of Heckman's Model. Panel data (2008-2019)**

| | | (3) Heckman | |
|---|---|---|---|
| Dep. Variable | Internet User | Coef. | z |
| Gender | Male | 0.016 | 0.810 |
| Age | 25-34 | **-1.124** | **-7.51** |
| | 35-44 | **-1.458** | **-11.41** |
| | 45-54 | **-1.772** | **-19.79** |
| | 55-64 | **-2.299** | **-44.87** |
| | 65-74 | **-2.848** | **-70.00** |
| | 75+ | -3.735 | . |
| Education | Secondary | **0.849** | **16.88** |
| | Bachelor's Degree | **1.666** | **13.870** |
| | Master or PhD | **1.774** | **13.340** |
| Digital Skills | Medium | **4.350** | **10.530** |
| | High | **2.859** | **19.130** |
| | Very High | **3.990** | **22.020** |
| Income | Medium | **0.535** | **13.740** |
| | High | **0.979** | **13.270** |
| | Very High | **1.024** | **11.380** |
| Population size | De 50.000 a 100.000 | **0.100** | **3.500** |
| | De 20.000 a 50.000 | **-0.061** | **-2.520** |
| | De 10.000 a 20.000 | **-0.257** | **-7.520** |
| | Less than 10.000 | **-0.428** | **-9.720** |
| | Constant | 0.065 | 0.340 |
| Observations | | 138,879 | |
| Selected | | 70,616 | |
| Non-selected | | 68,263 | |
| Groups | | 78,224 | |
| Wald $\chi^2$ | | 9,067.62; DF: 16 | |

Notes: Coefficients and z statistics significant at 5% are represented in bold. The z statistic is not reported for age 75+. Internet user in the last 3 months is the selection equation in Heckman's sample selection procedure (Equation 3). The adoption equation is reported in Table 3. Base categories are female, age 16-24, primary or no education, low digital skills, low income, and population size of 100,000+ and province capitals.