Differences in digitalization levels: a multivariate analysis studying the global digital divide

Billon, Margarita; Lera-Lopez, Fernando; Marco, Rocío

Rev World Econ (2010) 146:39–73
DOI 10.1007/s10290-009-0045-y

ORIGINAL PAPER

Differences in digitalization levels: a multivariate analysis studying the global digital divide

Margarita Billon · Fernando Lera-Lopez · Rocío Marco

Published online: 7 February 2010

Abstract This papers aims to identify and explain the differences in information and communications technologies (ICT) adoption for a sample of 142 developed and developing countries. In addition, we examine the relationships between specific combinations of technologies and the factors explaining them. Although income is a key factor for all country groups, its role is more significant for middle-digitalization countries. Using several multivariate techniques, we detect different patterns of digitalization. The patterns are explained to differing degrees by the type of country, by differences in economic development, and by socio-demographic and institutional variables. Factors such as quality of regulation and infrastructure explain ICT adoption in high-income countries. The ICT combination associated with specific income groups as well as the explanatory variables detected for each of them might be useful to implement the most appropriate policy actions to reduce the digital divide.

Keywords Digital divide · ICT · Digitalization · Internet · Economic development · Canonical correlation analysis

JEL Classification L86 · L96 · O33

M. Billon
Department of Economic Development, Autonomous University of Madrid, Carretera de Colmenar Viejo, Km 15, 28049 Madrid, Spain

F. Lera-Lopez (✉)
Department of Economics, Public University of Navarre, Campus Arrosadía s/n, 31006 Pamplona, Spain
e-mail: lera@unavarra.es

R. Marco
Department of Applied Economy, Autonomous University of Madrid, Carretera de Colmenar Viejo, Km 15, 28049 Madrid, Spain

1 Introduction

Given the fact that information and communication technologies (ICT) have been revealed to have remarkable impacts on economic development, the so-called digital divide has become an issue of great interest for researchers and policy makers. Disparities in ICT diffusion may lead to an increase in the disparities in terms of economic development. Thus, a large number of studies have focused on measuring and analyzing the nature of the digital divide.

Information and communications technologies diffusion has improved in many developing countries, particularly for some technologies, such as mobile phones or the Internet. Yet, the level of digital development is still much higher in the developed world (with some developing countries, such as Korea or China, being the exception). These days, the digital divide is increasingly related to differences in the speed and quality of access to ICT (ITU 2008).

The literature on the topic distinguishes between two main approaches: that focusing on measuring the gap for one specific technology or for a small group of countries, and that explaining ICT adoption. The latter usually refers to a single technology, such as personal computers, the Internet or broadband. Some studies elaborate upon an index grouping of technologies, although these frequently examine a small number of countries (Corrocher and Ordanini 2002; Bagchi 2005).

As different technologies show different patterns of diffusion (Rogers 2003; Karshenas and Stoneman 1995), the varied combinations of ICT may lead to diverse models of digitalization in different countries. Analysis of the digital divide should account for those differences. Therefore, the analysis of a single technology does not provide much information about the level of digital development within a country. A measure of digital development including several technologies would allow for comparisons between different levels of digitalization.

Within this framework, this paper's objective is twofold. First, we seek to identify and explain the differences in the digitalization levels between different groups of countries. Second, we aim to identify the relationships between specific combinations of technologies and the factors explaining them, mostly related to different development levels.

The literature has highlighted the role of income in explaining the adoption of some technologies, such as the Internet (Quibria et al. 2003; Beilock and Dimitrova 2003; Dewan et al. 2005; Chinn and Fairlie 2007), personal computers (Dewan et al. 2005; Chinn and Fairlie 2007), and broadband (Turk et al. 2008; Lee and Brown 2008). Nevertheless, some studies have also demonstrated the relevance of other non-economic factors, such as competition, telecommunication infrastructure and human capital (Quibria et al. 2003; Andonova 2006; Guillén and Suárez 2006; Oyelaran-Oyeyinka and Lal 2005). In the same vein, the differing combinations of ICT that shape diverse models of digitalization may be explained by a wide range of variables. These include income, as well as other institutional and non-economic factors pointing to a relationship between digitalization models and different levels of development.

Our study differs from those that create an index to measure the digital divide, such as Corrocher and Ordanini (2002), in that it includes different types of technologies capturing ICT use and infrastructure. Although Bagchi (2005) also creates an index to determine which factors contribute to the digital divide in various nations over time for four ICTs, the present study increases the number of technologies and extends the methodological approach. Along with principal component analysis and multiple regression analysis for each variable employed by other researchers (Chinn and Fairlie 2007; Dewan et al. 2005; Corrocher and Ordanini 2002; Bagchi 2005), we also use canonical correlation analysis. This approach allows us to detect different combinations of technologies and patterns of digitalization, as well as to explain them by several sets of variables. As far as we know, ours is the first study to use a single model to explain the digital divide and to capture its multidimensional nature. From a public policy perspective, the variables identified by the models can be useful to promote specific ICT measures according to the group of countries and the patterns of digitalization detected. Specific measures adapted to the characteristics of the digitalization patterns might be more suitable than global policies.

By including 142 developed and developing countries, we extend the geographical scope of other previous research considering a large number of technologies, such as Vicente and López (2006), whose study is related to the EU-15. We also extend the number of countries and technologies studied by Hargittai (1999) (18 OECD countries and the Internet) and the ICT covered by Chinn and Fairlie (2007) (computer and Internet use for the period 1999–2001).

The remainder of the paper is structured as follows. The following section presents the digital divide in ICT adoption. Section 3 provides the literature review. The research model and methodology are shown in Sect. 4. Section 5 describes the data and variables. Sections 6 and 7 present the analysis, models and findings. The final section presents the major conclusions and discusses issues for further research.

2 The digital divide in ICT adoption

The *digital divide* refers to “the gap between individuals, households, businesses and geographic areas at different socio-economic levels with regard to their opportunities to access information and communication technologies and to their use for a wide variety of activities” (OECD 2001, p. 5). Although there is general agreement on the definition of the digital divide, there is no common perspective to conceptualize and measure it (Vehovar et al. 2006). One of the reasons is the number and the variety of technologies involved. The digital divide can differ with the type of technology studied, since different technologies show different patterns of diffusion.

For a sample of 142 countries, Fig. 1 illustrates dispersion and inequalities in GDP and ICT adoption for several variables related to ICT use and infrastructure in 2004. We compute the digital divide using an inequality measure, such as the Gini index, and a dispersion measure, such as Pearson’s coefficient of variation.

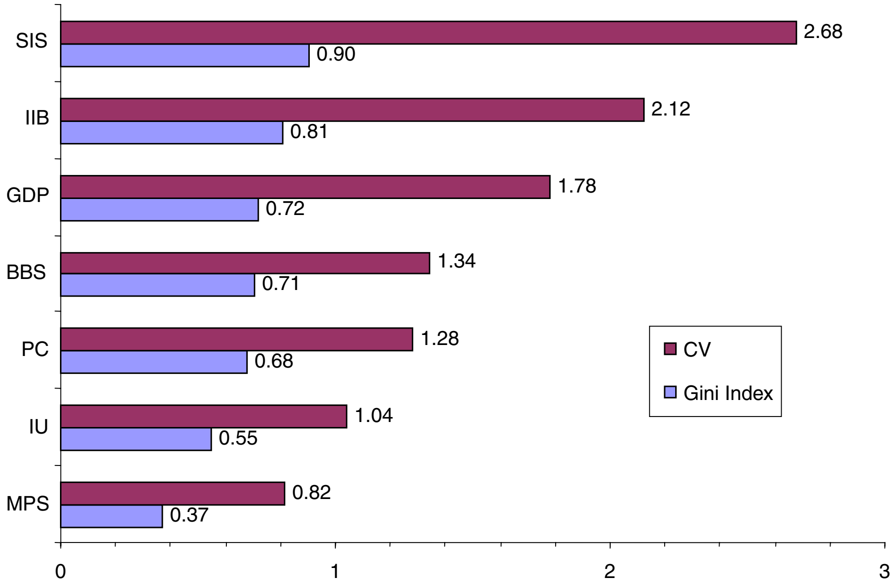

Fig. 1 Dispersion and inequalities between countries in ICT adoption. *Note*: Pearson's coefficient of variation is the ratio of standard deviation to the mean. Gini index is weightened by population. $N = 142$. *Source*: Author's work from World Bank (2006)

We show that mobile phone subscribers (MPS), Internet users (IU) and personal computers (PC) are more equally distributed than secure Internet servers (SIS) or broadband subscribers (BBS). As the literature on technology diffusion has found, some technologies are easier to be adopted than others. This is the case for mobile phones and the Internet, for example, which are easier for both firms and households to adopt relative to other technologies (James 2007).

Inequalities in ICT adoption may also depend on the different stages of the adoption process in which the countries are placed. The specific adoption pattern may also differ according to the different economic development levels. In Fig. 1, indicators associated with higher levels of infrastructure such as international Internet bandwidth (IIB) and the number of SIS, present the highest values for both Gini index and coefficient of variation. The inequality and dispersion in the adoption of these technologies are higher than those of GDP, while for the rest of technologies the inequalities are lower.

3 Literature review

From a theoretical standpoint, the methodological framework most frequently employed to explain diffusion rates of new technologies is diffusion theory (Rogers 2003). Adoption patterns are assumed to generally follow an S-shaped curve, with accelerated growth in the earlier stages of adoption and slower growth as technology matures (Rogers 2003; Karshenas and Stoneman 1995).

Some theoretical models explain ICT diffusion by accounting for the nature of the technology, the number and characteristics of the receiving users (consumers, households, firms, etc.), and the channels through which technology is transmitted. Adopters can be classified into different categories depending on the time they take to adopt an innovation. Diffusion rates also depend on the type of technology because some technologies are more likely to be adopted by some users than others. Consequently, different technologies are associated with different diffusion patterns.

According to epidemic models, the key to explaining ICT diffusion relies on the spread of information about the existence of a new technology and learning from experience (Mansfield 1968). The greater the number of adopters, the greater the probability that other users will be "contaminated", leading to the further spread of information and an accelerated diffusion speed (Karshenas and Stoneman 1995). Rank models add to the framework by considering heterogeneity and the characteristics of potential adopters in explaining diffusion. In turn, the literature on knowledge spillovers (Fujita and Mori 2005) contributes to diffusion theories by highlighting the role of the type of knowledge transmitted. When knowledge is tacit or non-codified, face-to-face communication facilitates its diffusion.

These contributions reveal a wide range of possibilities regarding the variables to include in empirical studies measuring and explaining ICT diffusion disparities between countries. The empirical literature about the digital divide can be divided into two groups. On the one hand, some studies focus on measuring and quantifying the digital divide, considering the evolution of the digital gap, in particular (OECD 2005; Corrocher and Ordanini 2002; Bagchi 2005; Vicente and López 2006). The multi-dimensional character of the digital divide has led to elaborate ICT indexes to summarize information about the level of digitalization, such as the Information Society Index, the Networked Readiness Index, the Digital Access Index, the Digital Opportunity Index and the Digital Divide Index (Vehovar et al. 2006). For instance, Corrocher and Ordanini (2002) combine six different dimensions of digitalization in an index to obtain several patterns of digitalization in ten developed countries. Bagchi (2005) creates an indicator including four different technologies, such as telephone (fixed and mobile), PC and Internet usage to measure and analyze the divide both globally and in different groups of countries. Vicente and López (2006) use ten variables, including PC, telephone lines, broadband connections and secure servers, to determine the digital divide between 15 European countries.

On the other hand, other empirical studies concentrate on explaining the determinants of ICT adoption and diffusion (Hargittai 1999; Kiiski and Pohjola 2002; Beilock and Dimitrova 2003; Dewan et al. 2005; Chinn and Fairlie 2007). Some researchers use cross-sectional data for a particular group of developed countries (Hargittai 1999; Vicente and López 2006), developing countries (Quibria et al. 2003; Wong 2002) or both (Beilock and Dimitrova 2003). Other studies extend the analysis to consider cross-sectional time series for developing countries (Tanner

2003; Oyelaran-Oyeyinka and Lal 2005; Dasgupta et al. 2005), while others include a combination of developing as well as developed countries.[1]

A summary of the most relevant studies on the digital divide is showed in Table 10 in the appendix. The common findings show the links between ICT diffusion disparities and GDP. Moreover, economic development seems to be a prerequisite of ICT diffusion as well as one of the most significant determinants of the digital divide. In fact, some authors consider the digital divide to be a new demonstration of the traditional differences in technological adoption between rich and poor countries (James 2007).

Despite the relevance of GDP in explaining the digital divide some studies highlight the fact that disparities in ICT adoption rates are greater than that of GDP (Wong 2002; Liu and San 2006). As showed in Fig. 1, the inequality and dispersion values are higher for some technologies than the values related to economic development. Thus, factors other than income may affect ICT diffusion. In fact, many researchers have highlighted the complex and multidimensional nature of the digital divide underlying the role of additional variables, such as educational, cultural, institutional, socio-demographic and political cross-country differences, to explain differences in ICT diffusion (Sciadas 2005; Corrocher and Ordanini 2002).

The likelihood of ICT diffusion is closely related to telecommunications infrastructure. Empirical evidence has shown that a greater level of ICT infrastructure seems to be associated with greater diffusion rates of some technologies (Quibria et al. 2003; Chinn and Fairlie 2007). Depending on the type of the study, a telecommunications infrastructure variable has been included as an explanatory variable or as a part of an index capturing the level of digitalization, as shown by Corrocher and Ordanini (2002).

The prices and the cost of access are usually found to be an additional influential factor for ICT diffusion. For a sample of 23 OECD countries, Kiiski and Pohjola (2002) show that a 50% reduction in the cost of Internet access would increase the number of computer hosts by 25% per capita over a 5-year period. Other authors have found that the cost of Internet usage has a negative impact on its usage (Demoussis and Giannakopoulos 2006). Nevertheless, other empirical evidence does not find a significant influence of telecommunications prices on ICT diffusion (Hargittai 1999; Andonova 2006; Chinn and Fairlie 2007).

Socio-demographic variables are also often cited as key factors for ICT diffusion. The role of education and the demographic features is particularly relevant (Hargittai 1999; Kiiski and Pohjola 2002; Quibria et al. 2003; Tanner 2003; Dewan et al. 2005). According to the diffusion theories mentioned previously (Rogers 2003), human capital is assumed to ease ICT diffusion because educated people will be more prone to adopting innovations such as the Internet (Quibria et al. 2003; Kiiski and Pohjola 2002; Crenshaw and Robison 2006). In addition, because the Internet is an interactive technology, specific skills often associated with high levels of education are needed to take advantage of ICT opportunities. Within the diffusion models mentioned above, population and its characteristics facilitate knowledge

[1] See Kiiski and Pohjola (2002), Dewan et al. (2005), Guillén and Suárez (2006), Bagchi (2005), Chinn and Fairlie (2007), Andonova and Diaz-Serrano (2009), and Pick and Azari (2008).

about new technologies. Demographic variables such as population size, population distribution and density and urban versus rural population are closely associated with the cross-country digital divide (Quibria et al. 2003; Dewan et al. 2005; Bagchi 2005; Chinn and Fairlie 2007).

Empirical studies support the notion that public policies and effective regulation are relevant factors in boosting or restricting ICT diffusion. Telecommunications policy may encourage ICT diffusion by developing new infrastructure, introducing more competition and reducing ICT access costs. Following the World Bank (2006), African and Latin American countries that introduced more competition into the telecommunications market during the 1980s and 1990s registered more accelerated growth in the ICT sector than the countries who postponed market competition. Also, Hargittai (1999) and Guillén and Suárez (2006) find that telecommunications policy has a significant influence on Internet adoption and use. Andonova and Diaz-Serrano (2009) show that the impact of political institution on ICT varies from one technology to another. This impact is higher in Internet use and fixed telephone use than in mobile phone use. Wallsten (2005) finds that regulation has a significant impact on Internet diffusion in developing countries. Dasgupta et al. (2005) emphasize the role of competition policy for developing nations to boost Internet use and mobile phone diffusion. However, findings from other studies show that this evidence might be ambiguous. For example, Kiiski and Pohjola (2002) point out that liberalization does not guarantee greater ICT diffusion alone. It must be accompanied by a reduction in prices. Other authors have emphasized the influence of regulation quality to explain the cross-country digital divide (Chinn and Fairlie 2007).

Empirical evidence on ICT adoption between developed and developing countries reveals remarkable differences in ICT diffusion patterns. Kiiski and Pohjola (2002) find that GDP per capita and the cost of Internet access are key factors in explaining ICT diffusion in OECD countries, while education is significant in explaining it in developing countries. Pick and Azari (2008) show that ICT diffusion is mainly associated with foreign investment and government prioritization of ICT for developing countries, rather than with educational and demographic variables. However, for developed countries, ICT diffusion is more heavily influenced by the participation of women in the labor force and by educational variables.

To sum up, the literature points to a wide set of economic, socio-demographic and institutional factors that may explain disparities in ICT diffusion within countries. The broad range of countries, technologies and variables involved in its diffusion reveal its multi-dimensional character and the complexity of the topic. In the next section, we present our research model as well as the methodology followed.

4 Research model and variables

Following the available empirical evidence and taking into account the multidimensional, relative and complex nature of the digital divide, our research model seeks to answer three main research questions and a set of varied hypotheses:

Research question 1: What are the factors explaining the determinants of ICT adoption in different groups of countries?

To answer the question we propose a model that estimates the relationship between the digitalization level, as measured by an index of digitalization, and a set of economic, socio-demographic and institutional factors for different groups of countries:

$$DigIndex = \beta_0 + \beta_1 GDP + \beta_2 TM + \beta_3 SLE + \beta_4 POP2 + \beta_5 IP20 + \beta_6 RQ + u \quad (1)$$

where the dependent variable *DigIndex* is an index capturing the level of digitalization for different groups of countries; *GDP* is GDP per capita; *TM* stands for telephone mainlines, *SLE* stands for school life expectancy; *POP*2 is the percentage of population between 15 and 64 years, *IP*20 is the Internet price for 20 h of use; *RQ* is regulatory quality and *u* is the error term. The digitalization index captures the following ICT indicators: PCs, IBB, SIS, IU, BBS and MPS.

Personal computers is a very common variable in many studies (Wong 2002; Quibria et al. 2003; Dewan et al. 2005; Bagchi 2005; Vicente and López 2006; Chinn and Fairlie 2007; Pick and Azari 2008) and it is usually related to Internet use. International Internet bandwidth refers to broadband infrastructure commonly used for the development of the Internet. The availability of advanced Internet protocol-based services would be impossible without the successful diffusion of broadband. There is growing empirical evidence on the determinants of broadband adoption between countries.[2] Secure Internet servers can be considered a proxy for the infrastructure needed for the development of e-commerce. This variable has been included in other studies as ICT infrastructure (Vicente and López 2006; Corrocher and Ordanini 2002).

Internet users have been widely used in many studies as the most important variable to describe ICT use (Wong 2002; Beilock and Dimitrova 2003; Tanner 2003; Oyelaran-Oyeyinka and Lal 2005; Bagchi 2005; Guillén and Suárez 2006; Chinn and Fairlie 2007). Given the accelerated growth in broadband diffusion and the detected differences among countries in its development, broadband subscribers have been incorporated into the analysis. Finally, we have included mobile phone subscribers (Quibria et al. 2003; Bagchi 2005; Pick and Azari 2008; Donner 2008) because mobile diffusion has dramatically increased in many countries.

Regarding our set of independent variables and considering the multidimensional character of the digital divide, we include four categories of factors: economic, socio-demographic, institutional and infrastructure.

As economic development seems to be a clear prerequisite for ICT diffusion, we have included GDP per capita following the empirical evidence. We expect a positive influence of GDP per capita on the digitalization index (Hargittai 1999; Kiiski and Pohjola 2002; Quibria et al. 2003; Dewan et al. 2005; Guillén and Suárez 2006; Bagchi 2005). This leads us to our first hypothesis:

H1. There is a positive and significant relationship between GDP and the digitalization index

[2] See Lee and Brown (2008) for a recent review of the literature.

The literature also shows that the likelihood of ICT diffusion is closely related to telecommunications infrastructure (Quibria et al. 2003; Chinn and Fairlie 2007). Consequently, the number of TM per 100 people has been incorporated as an explanatory variable. Following the prior literature we expect a positive influence of this variable on the digitalization index because it represents communication infrastructure available to access and use ICTs. The second hypothesis is the following:

H2. There is a positive and significant relationship between telephone mainlines and the digitalization index

We have employed SLE to measure the impact of education and POP2 to reflect the influence of age distribution on ICT adoption. We expect a positive influence of education on ICT diffusion given that more educated people are more prone to adopt innovations (Rogers 2003) and they may be also more qualified in terms of Internet skills. However, results from the prior literature are not conclusive: Kiiski and Pohjola (2002) find that education is not significant to explain Internet diffusion. Hargittai (1999) finds different results depending on the model. Quibria et al. (2003) show that the results depend on the technology and education level. For example, education is significant for computers and Internet use but not for cellular phones.

H3. There is a positive and significant relationship between education and the digitalization index

As mentioned above, according to epidemic models population plays a key role in technology diffusion. It has also been tested by the empirical evidence cited in the previous section. Evidence also shows that there is an inverse relationship between age and the adoption of some technologies. For example, Chinn and Fairlie (2007) include two different variables capturing the influence of population distribution on computer and Internet adoption: population between 0 and 15 years and population of 65 years and more. We have selected population between 15 and 64 to capture also the influence of population as well as the influence of age distribution. In general, we anticipate a positive sign for this variable because we expect it will capture the percentage of most active population in terms of ICT use.

H4. There is a positive and significant relationship between the percentage of population between 15 and 64 years of age and the digitalization index

Regarding institutional factors, we have considered the cost of Internet use (IP20) and the quality of regulation (RQ). This last variable measures the ability of the government to formulate and implement sound polices and regulations that permit and promote the development of the private sector. Following the empirical evidence, we expect the cost of Internet use to have a negative impact on ICT diffusion (Kiiski and Pohjola 2002; Vicente and López 2006) although other empirical evidence does not find a significant influence (Hargittai 1999; Andonova

2006; Chinn and Fairlie 2007). Regulatory quality might boost ICT diffusion by introducing competition into the telecommunications market (Wallsten 2005). Although it would contrast the results obtained by Chinn and Fairlie (2007) we expect the higher the regulatory quality, the higher the digitalization index.

H5. There is a negative and significant relationship between Internet price and the digitalization index

H6. There is a positive and significant relationship between regulatory quality and the digitalization index

As mentioned in previous sections, different technologies show different patterns of diffusion (Rogers 2003; Karshenas and Stoneman 1995). There are also important inequalities in ICT adoption according to the different economic development levels. Therefore, the varied combinations of ICT may lead to diverse patterns of digitalization in different countries. It leads us to propose the following research questions and hypothesis:

Research question 2: Are there different patterns of digital development defined as combinations of several ICTs?

Research question 3: What are the factors explaining the detected patterns of digital development?

To answer both questions we propose a model for measuring the relationships between the characteristics of the digital development and the set of explanatory variables included in the previous stage using canonical correlation analysis (CCA), a generalization of the usual linear regression model:

$$
\begin{aligned}
y_1^* &= \mathbf{Y}\beta = \beta_1 PC + \beta_2 IIB + \beta_3 SIS + \beta_4 IU + \beta_5 BBS + \beta_6 MPS \\
x_1^* &= \mathbf{X}\alpha = \alpha_1 GDP + \alpha_2 TM + \alpha_3 SLE + \alpha_4 POP2 + \alpha_5 IP20 \\
&\quad + \alpha_6 RQ + \alpha_7 LEVELDIG \\
&\quad \mathrm{CCA} \Rightarrow \max \mathrm{cor}\left(x_1^*; y_1^*\right)
\end{aligned}
\tag{2}
$$

where x^* and y^* are the canonical variables, linear combination of *X*'s and *Y*'s, respectively. Our set of dependent variables, $y_1,\ldots,y_q$, is the same as that included in the digitalization index: PCs, IBB,SIS,IU, BBS and MPS.

The set of independent variables is also the same set included in the regression model: *GDP*, *TM*, *SLE*, *POP*2, *IP*20, and *RQ*. In addition, the explanatory variables set includes a categorical variable (*LEVELDIG*) classifying the countries in three levels of digitalization: lowly, middle and highly digitalized countries.

Finally, the hypotheses to be tested in this model are the following:

H7. Different patterns of digitalization associated with different groups of countries can be detected

H8. Different patterns of digitalization are associated with economic development

H9. Different patterns of digitalization are explained by different types of demographic, social and institutional variables

The empirical evidence shown in Sect. 3 allows us to assume that other variables may be relevant to explain different patterns of digitalization. In this line, it should be mentioned that variables such as English proficiency (Hargittai 1999; Kiiski and Pohjola 2002; Chinn and Fairlie 2007), trade openness and foreign direct investment[3] were also selected in the initial models to be included in the CCA to capture the level of a country‘s integration in the global markets. English proficiency is also an education variable related to English skills to use the Internet. However, many authors cited above obtained that it is not significant. Following Chinn and Fairlie (2007), trade openness is related to the role of trade barriers and it may be considered a proxy for the level of regulation. However, neither English proficiency nor trade openness were significant in our initial model, so we decided to remove them to include other relevant variables. With respect to foreign direct investment could not be included in the analysis also because of the lack of data for most countries. We were also interested in including other variables such as ICT expenditure per capita (US$), telecommunications investment (% of GDP), population covered by mobile telephony (%), telecommunications revenue (% of GDP), expenditure per student (% of GDP per capita) and public education expenditure (% of GDP). However, we could not use them also because of the lack of data.

5 Methodology and data

In a first stage we seek to determine the factors explaining different levels of digitalization for different groups of countries. To capture the level of digitalization we begin by using a factorial analysis for the digitalization variables selected. We then create an index summarizing variables related to the access and use of several information and communication technologies.

We then use regression analysis to explain the digitalization index. However, regression analysis only allows us to analyze the influence of a set of variables on the digitalization indicator for different groups of countries studied separately. The model would only explain the variability captured by the index.

For this reason, we propose a model for measuring the relationships between the characteristics of the digital development and the set of explanatory variables used in the previous stage. We aim to determine whether the digitalization levels are related to different types of patterns of digitalization and to explain them. Finally, in an attempt to find a relationship between digitalization and development levels, we study whether the digitalization patterns are correlated with specific groups of countries.

To determine whether digitalization levels are related to different digitalization patterns and whether they are explained by those variables for different countries, we use CCA. Given the multidimensional nature of the digital divide and the variety of factors affecting it, CCA is a suitable technique that provides an additional contribution with respect to that of the multiple regression analysis, commonly used in other studies.

[3] See Hargittai (1999), Dewan et al. (2005), Crenshaw and Robison (2006), Chinn and Fairlie (2007), World Bank (2006), and Pick and Azari (2008).

In the case of multiple regression analysis, when the dependent variable is a synthetic index (the first factor obtained by factor analysis) the technique only allows us to explain the common information of the elementary variables included in the index. This common information is the proportion of the total variability captured by the first factor. In contrast, CCA allows us to explain the total variability of the set of the representative variables of digitalization.

Canonical correlation analysis seeks to identify and quantify the association between two groups of variables (Johnson and Wichern 2007). In our case, these sets are the digitalization or dependent variables set (Y) and the explanatory variables set (X) mentioned above. CCA translates the relationships between (and across) the two sets of variables into a parsimonious number of linear combination of variables with the greatest correlations, which summarize the entire variable space.

In CCA, linear combinations y^* and x^* provide simple summary measures of the set of digitalization variables Y and the explanatory variables X. Set:

$$\begin{aligned} y^* &= \mathbf{Y}\beta = \beta_1 y_1 + \beta_2 y_2 + \ldots + \beta_q y_q \\ x^* &= \mathbf{X}\alpha = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_p x_p \end{aligned} \tag{3}$$

for the same pair of coefficient vectors $\boldsymbol{\alpha}$ y $\boldsymbol{\beta}$. We seek coefficient vectors $\boldsymbol{\alpha}$ y $\boldsymbol{\beta}$ such that the canonical correlation between linear combinations is as large as possible:

$$\mathrm{Corr}(x^*, y^*) = \frac{E[\alpha' x y' \beta]}{E[\alpha' x x' \alpha]^{1/2} E[\beta' y y' \beta]^{1/2}} = \frac{\alpha' V_{12} \beta}{(\alpha' V_{11} \alpha)^{1/2} (\beta' V_{22} \beta)^{1/2}} = \rho \tag{4}$$

$\boldsymbol{V}_{11}$ = Cov ($\boldsymbol{X}$), $\boldsymbol{V}_{22}$ = Cov ($\boldsymbol{Y}$) and $\boldsymbol{V}_{12}$ = Cov ($\boldsymbol{X}$,$\boldsymbol{Y}$) are the covariance matrices. The first pair of canonical variables is the pair of linear combinations x_1^*, y_1^* with unit variances that maximize the correlation in Eq. 4. The second pair of canonical variables is the pair of linear combinations x_2^*, y_2^* with unit variances that maximize correlation 4 among all choices that are uncorrelated with the first canonical pair, and so on. The maximization aspect of the technique attempts to summarize the high-dimensional relationship between the two sets of variables into a few pairs of canonical variables, which are easier to be interpreted.

Regard to the data, the final database includes 142 countries for the year 2004.[4] The dichotomized correlations test (Tabachnick and Fidell 1996) verifies the randomness in the generation of missing values, which guarantees our ability to generalize the results. The sample covers 98.8% of the world's total population and includes 48 low-income, 65 middle-income and 29 high-income countries, according to World Bank's classification.

Most of the digitalization variables are from the World Telecommunication Development Database (ITU 2006). Economic, socio-demographic and Internet price variables are from the World Development Indicators Database (World Bank) while the variable indicating regulatory quality is from the Worldwide Bank's

[4] Out of the 208 countries included in the World Development Indicators Database in December 2007, we excluded those with total population lower than 1 million inhabitants in 2004 (56 countries) and those with lack of data in 3 or more of the 12 variables considered in the analysis (10 countries). The final database is composed by 1,657 observations, 47 missing.

Governance Indicators Database (World Bank). Table 1 contains a brief description of the variables, their sources and the main descriptive statistics.

The exploratory analysis obtained from the main descriptive statistics shows a distribution that is far from normal. It is characterized by high ratios of outliers, positive and high coefficients of skewness and kurtosis for many variables. To agree with the assumption of normality supposed in a multiple regression analysis, we transform the original variables. The logarithmic transformation of the original data greatly improves the appearance of normality and does not reduce the interpretative power of the model.[5] The logarithmic transformation also improves the linearity of the relationships between variables, which is another advantage of the technique, since the assumption of linearity is required for both canonical correlation and multiple regression analysis.

6 Digitalization and development levels: bivariate analysis

As mentioned above, we begin our analysis by measuring the level of digitalization that will allow us to compare ICT development between countries. Given that the purpose of this bivariate analysis is mainly descriptive, we use the original variables for an easier interpretation of the relationships between digitalization and development. In this first step, we are interested in creating a digitalization index from our selected digitalization variables: BBS, IU, SIS, PCs, IIB, and MPS. We run the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and the Barlett test of sphericity to test whether we can employ factor analysis.[6] The value of 0.82 in KMO measure and the probability lower than 1% associated to the Barlett test value suggest that the data structure is adequate for factor analysis (Table 2). Our factor extraction relied on principal component analysis. By eigenvalue criterion, we finally only consider one factor (with an eigenvalue greater than 1). The high communality among the digital variables allows us to explain 77.8% of the total variance of the selected variables set with the first factor, our digitalization index.

Each country's digitalization level was measured by multiplying the factor score coefficients of each variable by their standardized values. We can only obtain score for the 116 countries with available data for the six ICT variables.

Figure 2 shows the relationship between GDP per capita and the digitalization index obtained for each country. For most of the countries considered in the analysis, there is a positive relationship between the two variables, confirming the available empirical evidence. Variability rises as the value of the variables increases.

The positive linear relationship is corroborated by a high value of Pearson's correlation coefficient ($r = 0.914$). We expect a high similarity between the country

[5] With the transformation to the logarithmic scale, the problem of outliers in the data disappears and the transformed variables become symmetric and mesokurtic to a large extent. These transformations are consistent with those carried out in the literature and show the non-normal shape of the data.

[6] KMO measure requires values greater than 0.5 for running a factor analysis. Barlett tests the null hypothesis that the correlation matrix is an identity matrix, which implies factor analysis would not be suitable.

Table 1 Variables, main descriptive statistics and sources

Variable	n	Minimum	Maximum	Mean	Median	SD	Description	Source
BBS	139	0.000	24.794	2.672	0.069	5.387	Broadband subscribers (per 100 people)	International Telecommunication Union, World Telecommunication Development Report and database, and World Bank estimates
IU	142	0.080	75.620	15.833	7.715	18.795	Internet users (per 100 people)	International Telecommunication Union, World Telecommunication Development Report and database, and World Bank estimates
MPS	142	0.210	121.000	36.319	25.695	34.168	Mobile phone subscribers (per 100 people)	International Telecommunication Union, World Telecommunication Development Report and database, and World Bank estimates
PC	139	0.020	82.620	14.669	4.761	21.335	Personal computers (per 100 people)	International Telecommunication Union, World Telecommunication Development Report and database, and World Bank estimates
IIB	141	0.040	34,891.47	1,406.56	33.290	4,226.55	International Internet bandwidth (bits per person)	International Telecommunication Union, World Telecommunication Development Report and database, and World Bank estimates
SIS	122	0.010	674.630	52.702	3.714	116.120	Secure Internet servers (per 1 million people)	Netcraft (http://www.netcraft.com/)
TM	142	0.020	71.090	18.736	12.271	19.084	Telephone mainlines (per 100 people)	International Telecommunication Union, World Telecommunication Development Report and database, and World Bank estimates

Table 1 continued

Variable	n	Minimum	Maximum	Mean	Median	SD	Description	Source
IP20	142	3.850	167.510	37.111	27.870	30.731	Internet price, 20 h of use (US$) August 2004	World Telecommunication/ICT Development Report 2006
GDP	141	86.450	39,804.81	6,350.14	1,707.28	9,675.16	GDP per capita (constant 2000 US$)	World Bank national accounts data
POP2	142	48.020	78.560	61.937	63.409	6.792	Population ages 15–64 (% of total)	World Bank
SLE	122	3.480	20.360	12.199	12.286	3.267	School life expectancy (years), total	World Bank, EdStats
RQ	142	0.280	4.370	2.506	2.313	0.945	Regulatory quality	World Bank, Worldwide Governance Indicators database

Source: Author's own calculations

Table 2 Digitalization index: factor analysis results

Factor	Eigen-value	Total variance explained		Variables	Factor 1	
		Percent of variance	Cumulative percent of variance		Factor loadings	Communality
1	4.667	77.79	77.79	PC	0.973	0.947
2	0.542	9.03	86.82	IIB	0.770	0.594
3	0.447	7.45	94.26	SIS	0.834	0.696
4	0.213	3.55	97.81	IU	0.941	0.886
5	0.087	1.45	99.26	BBS	0.918	0.843
6	0.044	0.74	100	MPS	0.838	0.702
KMO measure of sampling adequacy			0.815			
Barlett test of sphericity			804.1 [$p = 0.000$]			

Extraction method by principal component analysis

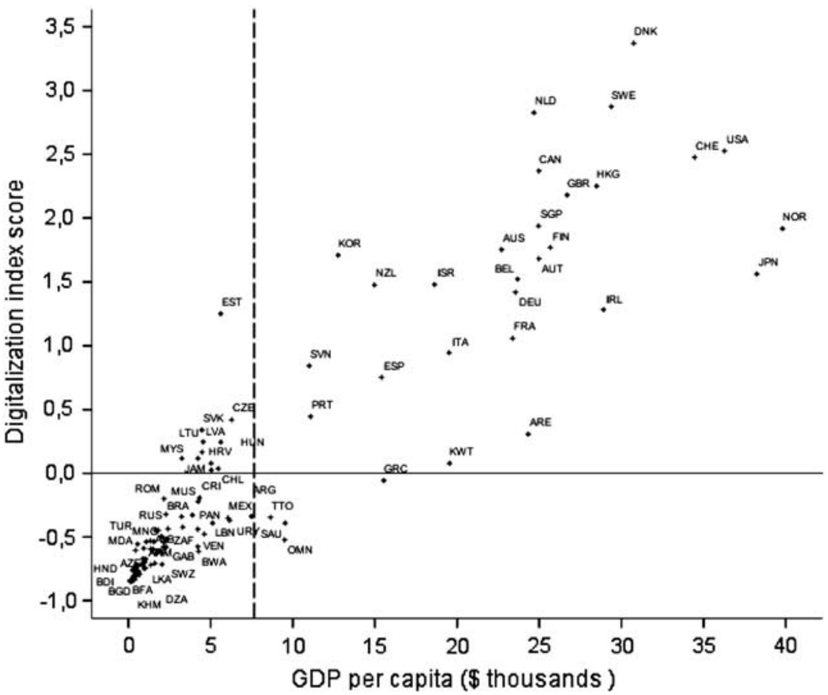

Fig. 2 GDP per capita versus digitalization index. *Source*: Author's own calculations

rankings by income and digitalization. The significant Spearman's correlation coefficient (value of 0.948) confirms the narrow relationship between GDP and the digitalization level. Mean values of GDP and the digitalization index divides the figure into four areas. The upper right quadrant (second quadrant) includes mainly OECD countries with GDP and digital levels higher than the mean. The lower left quadrant (the third quadrant) shows countries with lower GDP and lower levels of digitalization and includes Asian countries (e.g., Cambodia and Sri Lanka), Latin American countries (e.g., Brazil, Mexico, and Venezuela) and some Eastern countries (Romania, Russia, and Moldova). As we can see, few countries are located in the first and fourth quadrants (cases with inverse relationships between GDP and the digitalization level). The first quadrant includes Eastern economies registering a GDP level lower than the average but a digitalization index higher than that of most developing countries (Estonia, Czech Republic, Lithuania, and the Slovak Republic). Finally, the fourth quadrant comprises countries with an above average GDP level but with a below average digitalization index. Countries such as Trinidad and Tobago, Oman and Saudi Arabia are located in this quadrant.

We identified possible bivariate outliers using the Mahalanobis distance measure. Six countries are significantly far from the center of gravity or mean vector. On the one hand, Japan, Norway and the United Arab Emirates present a low-digitalization level compared to their high levels of GDP. On the other hand, Estonia, the Netherlands and Denmark show a high digitalization level given the level of their GDP per capita.

After creating the digitalization index and comparing it with the level of GDP, we aim to use regression analysis to explain the digitalization level by a variety of economic, socio-demographic and institutional variables.

7 Models and findings

7.1 Factor and regression analysis

Due to the regression model assumptions, we work with the log transformed variables of the original data, so we will obtain the elasticity coefficients index-predictors. The dependent variable is the digitalization index. For consistency, we have created a new digitalization index with the log transformed ICT variables. As with the original data, the KMO measurement and Barlett test confirms that factor analysis is a suitable technique. The results of the factor analysis are shown in Table 11 in the appendix. After extracting by principal component method, we obtain one relevant factor which explains 88.26% of the total variance.[7] The use of the new digitalization index-log as dependent variable does not alter the results of the proposed regression model.[8]

We divide the sample of countries into thirds, according to the digitalization index score obtained. The three groups are shown in Table 12 in the appendix. They are related to specific income groups. Following the World Bank's classification, the first one includes mostly OECD countries (with Eastern countries, Hong-Kong, Kuwait and South Korea as exceptions). The second group consists of mainly Latin

[7] The gain of explained variance with respect to the former digitalization index means a linear improvement of the relationships due to the logarithmic transformation.

[8] The ranking of countries according the digitalization index with the original variables and that with the log variables is practically the same, as shown by the Spearman's coefficient value (0.99) between both indexes.

Table 3 Regression result for the digitalization index

	High digitalization		Middle digitalization		Low digitalization	
	β	Stand. β	β	Stand. β	β	Stand. β
GDP	0.104** (0.046)	0.284	0.185*** (0.050)	0.515	0.230** (0.091)	0.395
TM	0.223** (0.101)	0.301	0.020 (0.073)	0.040	0.238*** (0.071)	0.649
SLE	−0.165 (0.290)	−0.069	0.701** (0.320)	0.265	0.056 (0.254)	0.031
POP2	−1.220** (0.495)	−0.201	1.147* (0.635)	0.273	−0.314 (0.963)	−0.059
IP20	−0.025 (0.045)	−0.044	0.106 (0.070)	0.207	0.097 (0.083)	0.152
RQ	1.159*** (0.280)	0.493	0.159 (0.180)	0.121	−0.106 (0.125)	−0.089
F-test	25.351***		8.629***		13.207***	
Adjusted R^2	0.802		0.574		0.709	
Sample size	37		35		31	

The dependent variable is the index of digitalization. Variables are log transformed. Values are unstandardized and standardized beta coefficients from OLS regressions. Standard errors in parenthesis

* Significant at 10% level ($p < 0.1$)

** Significant at 5% level ($p < 0.05$)

*** Significant at 1% level ($p < 0.01$)

American and Asian middle-income countries. Finally, the last group consists of low-income countries.

We run regression analysis for each of the three digitalization groups. The dependent variable is the digitalization index obtained for each group. As independent variable we include GDP per capita, the number of telephone mainlines lines (TM) to capture the role of infrastructure, school life expectancy (years) (SLE) representing education level, the percentage of the population between 15 and 64 years of age (POP2), Internet price (IP20) and regulatory quality (RQ) as institutional variables.

Table 3 presents the OLS cross-section estimation results from the regression analysis, showing the differing relevance of the explanatory variables in each digitalization group. The global significance *F*-test and the adjusted R^2 obtained describe the fit of the regressions.

The estimates imply that the partial elasticity of the digitalization index respect to GDP per capita is 0.104 for high-digitalization countries. This means that, for an increase of 1% in the GDP, *ceteris paribus*, the digitalization index will increase by 0.104%. As shown by the adjusted R^2, the digitalization index is better explained in the extremes: countries with high or low levels of digitalization. The relationship between the explanatory variables and the digitalization index is weaker for middle-digitalization countries.

A first conclusion to be drawn from the regression analysis is that different digitalization levels are explained by different variables. The role of GDP is more remarkable for middle-income countries, low-income countries and high-income countries, respectively. As the standardized coefficients in Table 3 show, the quality of regulation has the greatest positive influence on the dependent variable for the highly digitalized countries, followed by infrastructure and GDP and the negative influence of the population percentage between 15 and 64 years. As mentioned, we expected a positive sign for this variable. However, the variable shows a negative sign for highly digitalized countries and positive sign for middle-digitalization countries. This fact may be explained by the differences in age distribution between the developed and developing world as well as by ageing in highly digitalized countries. In addition, the available variable, as provided by WDI database, captures a wide range of ages (15–64 years). ICT use, for example Internet use, may vary substantially among this range of ages.

Infrastructure is by far the most important variable for the lowly digitalized countries, followed by GDP. Finally, the main determinant is GDP among the middle-digitalization countries, followed by the population between 15 and 64 years and education.

7.2 Canonical correlation analysis

After identifying the variables explaining the digital development for each group, we seek to explain the relationships between the specific technologies used to create the index and the set of explanatory variables. We are interested in knowing whether we can identify different types of patterns of digitalization as well as the variables explaining them.

However, the regression analysis cannot allow us to detect the existence of different patterns of digitalization. We only explain the common variability of the digitalization variables captured by the factor. The scenario where the single factor obtained captures an important share of the variance of the dependent variables set is valid, but can be improved. CCA allows us to explain all the information within the dependent variable set, including the non-common variability undetected by the index factor.

Model 1. Our dependent variables are those included in the digitalization index (PC, IIB, SIS, IU, BBS and MPS). The set of independent variables is the same as that included in the regression models. In this case, we also incorporate dummy variables indicating the digitalization level according to the three digitalization groups (high, middle or low). We seek to identify the group of countries related to the digitalization patterns detected.

Table 4 presents the correlation matrix between the digitalization variables Y and the explanatory set X. As we can see, the linearly assumption needed for CCA is more than accomplished. The correlations between GDP per capita and all the digitalization variables are remarkable, as happens for most of the independent variables. However, the price of 20 h of Internet use (IP20) is negatively associated with all the digitalization variables, showing the weakest correlation with SIS (−0.195).

The CCA results in Table 5 include a battery of four multivariate statistics testing the overall model fit. The null hypothesis (that the two sets of variables are not linearly related) is rejected at α level 0.05 in all four multivariate statistics. Before interpreting the canonical variates and canonical correlations, we need to determine the number of significant dimensions. Statistical significance is tested by computing

Table 4 Pearson correlations between *X* and *Y* sets

Rxy	BBS	IU	MPS	PC	IIB	SIS
IP20	−0.443**	−0.504**	−0.367**	−0.480**	−0.373**	−0.195*
GDP	0.835**	0.876**	0.874**	0.911**	0.910**	0.915**
SLE	0.706**	0.804**	0.784**	0.833**	0.770**	0.753**
POP2	0.744**	0.788**	0.734**	0.775**	0.738**	0.608**
TM	0.799**	0.904**	0.828**	0.894**	0.847**	0.818**
RQ	0.669**	0.677**	0.749**	0.714**	0.775**	0.760**

Pairwise Pearson correlation (*N* maximum 142, *N* minimum 122). Variables log transformed

* Correlation is significant at 0.05 (2-tailed)

** Correlation is significant at 0.01 level (2-tailed)

the Chi-square sequential test statistic.[9] Using the Chi-squared test, we find that the first four canonical correlations appear to be non-zero (at the same 0.05 α). However, as explained by Johnson and Wichern (2007) and Hair et al. (1998), we cannot rely on statistical significance tests to determine the number of significant dimensions. Redundancy analysis is also required to test the practical significance of canonical correlations.

From a practical viewpoint, redundancy analysis shows that the third and subsequent samples of canonical correlations can be ignored because they are comparatively smaller in magnitude and the corresponding canonical variates explain very little of their own variation. The total variance of the digitalization set explained by the independent set (total redundancy index) is 84.5%, but the explained variance is concentrated in the two-first canonical variates (70.8% + 13.0% = 83.8%). As a result, the first two dimensions with practical significance are the relevant canonical variables to be considered for the interpretation of the model.

Table 6 shows the canonical loadings and the canonical coefficients for both sets of variables. The canonical loading shows the correlation between the canonical variates and the original variables and they provide only bivariate information. The canonical coefficient quantifies the variable effect, taking the effect of the remainder of the variables in the model into account. Therefore, loadings and coefficients may have different signs. In both cases, the largest values (in absolute terms) are used to interpret the results.

As mentioned, the first canonical variate pair explains 70.8% of the variability of the dependent set. Although all variables are positively related to the digitalization indicator, y_1^*, PC (loading 0.902) and IU (loading 0.906) show the highest weight. Given the high level of the canonical loadings and coefficients for IIB and BBS, this dimension presents a digitalization pattern that may be characterized by a high relative share of those variables related to Internet use. Although SIS is positively

[9] The null and alternative hypotheses for assessing the statistical significance of the first *k*th canonical correlations are $H_0^k: \rho_1 \neq 0,\ \rho_2 \neq 0,\ \ldots,\ \rho_k \neq 0,\ \rho_{k+1} = 0,\ \ldots,\ \rho_p = 0 \quad H_1^k: \rho_i \neq 0,\quad \text{for same } i \geq k+1$ which has an approximate Chi-square distribution assuming multivariate normal data.

Table 5 Canonical correlations

Multivariate statistics	Value	*F*-value	*p*-value	
Measures of overall model fit				
Wilks' lambda	0.002	41.825	0.000	
Pillai's trace	2.255	11.543	0.000	
Hotelling-Lawley trace	53.614	162.543	0.000	
Roy's greatest root	45.453	870.106	0.000	
Canonical pair	**Canonical correlation**	**Chi-square test**	***df***	***p*-value**
Canonical correlation test				
1	0.989	849.40	48	0.000
2	0.927	335.65	35	0.000
3	0.447	60.75	24	0.000
4	0.350	30.51	15	0.010
5	0.276	12.84	8	0.117
6	0.119	2.07	3	0.558
Canonical pair	**Squared correlation**	**Variance extracted in set *Y***	**Redundancy measure y_i^*/x_i^***	
Canonical redundancy analysis				
1	0.978	0.724	0.708	
2	0.859	0.148	0.130	
3	0.200	0.022	0.004	
4	0.123	0.017	0.002	
5	0.076	0.003	0.000	
6	0.014	0.003	0.000	
Total redundancy *Y/X*			0.845	

related to the first digitalization indicator (in bivariate terms, positive loading), its negative coefficient in a multivariate framework points to countries with relatively lower level of SIS.

With regard to the set of independent variables, the first digitalization indicator, y_1^*, is mainly and positively explained by the percentage of the population between 15 and 64 years (POP2) and infrastructure (TM), to a lesser extent by education (SLE) and GDP, and negatively by Internet prices (IP20). We obtain the expected signs. These results are also in line with those of other, previously mentioned studies where Internet use, for example, is explained by demographic factors, some socio-economic variables such as education, and Internet prices.[10] The dummies included indicate that this dimension refers to highly digitalized countries.

[10] See Kiiski and Pohjola (2002), Corrocher and Ordanini (2002), Dewan et al. (2005), Guillén and Suárez (2006), Oyelaran-Oyeyinka and Lal (2005), and Andonova (2006).

Table 6 Canonical variables

Set Y	y_1^*		y_2^*	
	Canonical coefficients	Canonical loadings	Canonical coefficients	Canonical loadings
PC	0.737	0.902	−0.884	0.288
IIB	0.339	0.867	−0.168	0.382
SIS	−0.846	0.741	2.233	0.626
IU	0.305	0.906	−0.720	0.231
BBS	0.307	0.860	−0.423	0.235
MPS	0.120	0.817	0.374	0.401
Set X	x_1^*		x_2^*	
	Canonical coefficients	Canonical loadings	Canonical coefficients	Canonical loadings
GDP	0.005	0.843	1.035	0.458
TM	−0.325	0.868	0.989	0.271
SLE	0.230	0.840	−0.316	0.146
POP2	0.376	0.876	−0.648	0.000
IP20	−0.446	−0.657	0.924	0.507
RQ	−0.032	0.655	0.304	0.478
HIGHDIG	0.750	0.708	−0.689	0.377
MIDDIG	0.309	0.041	−0.155	−0.058

Variables are log transformed. Canonical coefficients are standardized

Although the number of telephone mainlines lines is positively related to the digitalization indicator in bivariate terms (positive loadings), its contribution is negative (−0.325) in a multivariate framework. This means that the greater the infrastructure, the greater the y_1^*. However, this points to countries with relatively lower TM when we account for the rest of the variables.

In this pattern, the greater the GDP, the greater the Internet use, although the impact of GDP is practically null when the rest of independent variables are accounted for in a multivariate framework.

Given the information provided by the canonical coefficients and loadings, the first dimension points to a digitalization model associated with highly digitalized countries, with a high proportion of adults in the population, education, income and infrastructure, and with a moderate role of Internet prices and regulatory quality.

The second dimension explains an additional 13% of the variability of the dependent set. The canonical loadings and coefficients show a digitalization pattern characterized by the role of SIS and, to a lesser extent, by MPS. In comparison with the results obtained for the first dimension, other variables such as IU and BBS are not relevant.

This pattern is positively explained by IP20 and GDP and, to a lesser extent, by RQ. This is surprising, since we would have expected a negative sign for Internet costs, given that secure Internet servers can be considered as a proxy for e-commerce and its correlation with IP20 is negative although weak (see Table 4).

However, as in the first dimension, IP20 has a negative influence on digital diffusion. In the second dimension, this variable could be a proxy for the positive influence of quality improvements and better innovative telecommunication services in e-commerce diffusion. Also, as the second dimension captures the variability not captured by the first dimension, the role of IP20 in this case may be related to countries with high GDP and also high Internet prices. As we previously mentioned, it should be noted that the empirical evidence regarding the sign of Internet price is not conclusive. Along with IP20, the second dimension is explained by GDP and, to a lesser extent, by RQ.

As with the first dimension, the canonical loadings show that the second pattern is also related to highly digitalized countries. This is an interesting result. The relationships captured by CCA between the two sets of variables, which have been grouped in two independent dimensions, are both related to countries with a high level of digitalization. However, the results also show that two different patterns are explained by different independent variables within the highly digitalized group. Countries following the first digital diffusion pattern are Korea, Hong-Kong, the Slovak Republic and Estonia. Countries included in the second digital diffusion pattern are, for example, the United Kingdom and Switzerland.

The relationships between the two sets of variables are stronger for the highly digitalized countries. Thus, they probably hide the relationships that might be found for the rest of the countries. To overcome this shortcoming, we proceed to carry out the CCA for the middle and lowly digitalized economies.

Model 2. The analysis of CCA for the whole sample provides two dimensions both related to highly digitalized countries. It may be interpreted as a sign of the extent of the digital divide between developed and developing countries. However, we are also interested in detecting digitalization patterns for the other groups of countries: middle and lowly digitalized economies. For this reason, in the Model 2 we run the CCA only including the 77 countries not belonging to the highly digitalized group. We consider the same dependent and independent variables set as in the Model 1. The variable capturing the digitalization level has two categories in this case: middle and lowly digitalized groups. The inclusion of the dummy (MIDDIG) seeks to test whether the CCA dimensions can be associated with middle or lowly digitalized countries.

The overall fit of the model confirms the existence of a linear relationship between the two sets of variables, with two canonical correlations statistically different from zero at 0.05 α level (Table 7). The redundancy measures inform us that only the first pair of canonical variates has practical relevance, accounting for 68.2% of the variance of the dependent set. The lower redundancy level indicates that the linear relation between the two variables set is weaker in this sample related to middle and lowly digitalized countries, in which only one pattern of digitalization has been identified.

Table 8 shows that all ICT variables have a positive and high association (canonical loadings) with the digital indicator y_1^*, BBS being the lowest. However, considering both the canonical coefficients and loadings, the dimension identified reveals that PC, IIB and MPS are the main variables of this combination. We might conclude that the digitalization pattern is less defined in comparison to the first

Table 7 Canonical correlations

Multivariate statistics	Value	*F*-value	*p*-value	
Measures of overall model fit				
Wilks' lambda	0.039	7.238	0.000	
Pillai's trace	1.665	3.785	0.000	
Hotelling-Lawley trace	10.521	15.615	0.000	
Roy's greatest root	9.400	92.657	0.000	
Canonical pair	**Canonical correlation**	**Chi-Square Test**	***df***	***p*-value**
Canonical correlation test				
1	0.951	224.53	42	0.000
2	0.647	62.94	30	0.000
3	0.431	25.50	20	0.183
4	0.341	11.33	12	0.501
5	0.179	2.824	6	0.831
6	0.092	0.590	2	0.745
Canonical pair	**Squared correlation**	**Variance extracted in set *Y***	**Redundancy measure y_i^*/x_i^***	
Canonical redundancy analysis				
1	0.904	0.755	0.682	
2	0.419	0.049	0.021	
3	0.186	0.051	0.009	
4	0.116	0.061	0.007	
5	0.032	0.048	0.002	
6	0.008	0.036	0.000	
Total redundancy *Y/X*			0.721	

Middle- and low-digitalization countries

dimension obtained in Model 1, given that it combines a variable showing mobile phones use with those that can be defined as technologies usually highly correlated to the Internet (PC and IIB).

These variables are mostly explained by the GDP and the number of telephone mainlines. Although canonical loadings for high education (SLE) and population (POP2) are also high, their contributions in multivariate terms are not relevant. The loading for MIDDIG confirms that this pattern is related to countries classified as middle-digitalized countries.[11]

[11] As the digitalization level is a dichotomous variable (middle/low country), the coefficients and loadings would be the same if we introduce the LOW-DIG dummy, but would have opposite signs. A −0.861 canonical loading for LOW-DIG means that the countries captured by that dimension are not lowly digitalized countries.

Table 8 Canonical variables

Set Y	y_1^*	
	Canonical coefficients	Canonical loadings
PC	0.321	0.884
IIB	0.279	0.912
SIS	0.145	0.879
IU	0.088	0.899
BBS	0.050	0.753
MPS	0.248	0.876
Set X	x_1^*	
	Canonical coefficients	Canonical loadings
GDP	0.398	0.933
TM	0.396	0.902
SLE	0.066	0.760
POP2	−0.033	0.712
IP20	0.070	−0.372
RQ	0.032	0.457
MIDDIG	0.298	0.861

Variables are log transformed. Canonical coefficients standardized

Middle- and low-digitalization countries

8 Discussion and conclusions

In this paper we have answered several important research questions and tested several hypotheses related to the digital divide. A summary of the main results is shown in Table 9. Regarding our first question "*What are the factors explaining the determinants of ICT adoption in different groups of countries?*" we have found that the variables explaining the level of digitalization in 142 countries differ according to the group of countries considered. From a digitalization index created by factor analysis and after applying OLS cross-section analysis, we have found that GDP is the only factor that seems to have a significant effect on digital development for all country groups. Nevertheless, its relevance varies according to the group. The results show that its role is more significant for middle digitalized countries and less relevant for highly digitalized countries where other non-economic variable become more important (Hypothesis 1). In line with the prior literature, the relationship between GDP and the digitalization index is positive for the three groups.

Our analysis also shows that there is a positive and significant relationship between telephone mainlines and the digitalization index for lowly and highly digitalized countries (Hypothesis 2).

We have also found that there is a positive and significant relationship between education and the digitalization index for middle digitalized countries (Hypothesis 3).

Table 9 Summary of results

		Expected sign	Results
H1. There is a positive and significant relationship between GDP and the digitalization index	Supported for highly and lowly digitalized countries. Strongly supported for middle digitalized countries	(+)	(+)
H2. There is a positive and significant relationship between telephone mainlines and the digitalization index	Strongly supported for lowly digitalized countries. Supported for highly digitalized countries	(+)	(+)
H3. There is a positive and significant relationship between education and the digitalization index	Supported for middle digitalized countries	(+)	(+)
H4. There is a positive and significant relationship between the percentage of population between 15 and 64 years of age and the digitalization index	Weakly supported for middle digitalized countries Refuted for highly digitalized countries	(+)	(+) for middle digitalized countries (−) for highly digitalized countries
H5. There is a negative and significant relationship between Internet price and the digitalization index	Refuted for the three groups	(−)	Not significant
H6. There is a positive and significant relationship between regulatory quality and the digitalization index	Strongly supported for highly digitalized countries	(+)	(+)
Canonical correlation analysis			
H7. Different patterns of digitalization can be detected for different groups of countries	Strongly supported for highly digitalized countries. Partially supported for middle digitalized countries		
H8. Different patterns of digitalization are associated with different levels of economic development	Strongly supported for highly digitalized countries (Model 1, CCA) and partially supported for middle digitalized countries (Model 2, CCA)		
H9. Different patterns of digitalization are explained by different types of economic, social and institutional variables	Supported		

With respect to the positive and significant influence of the population between 15 and 64 years on the digitalization index (Hypothesis 4) we have found a positive relationship for middle digitalized countries but a negative and significant influence for highly digitalized countries. This finding may be related to differences in age distribution between developed and developing countries. It may be also associated with ageing in highly digitalized countries and also to the fact that the variable includes a wide range of ages in which differences in ICT adoption may be considerable.

The hypothesis that there is a negative and significant relationship between Internet price and the digitalization index has been refuted for the three groups of countries (Hypothesis 5). It should be noted that empirical evidence in relation to the sign is not conclusive. First, we consider that this may be explained by the fact that in the regression analysis we use an index capturing several technologies, not only the Internet use. In addition, the results show that other variables related to economic development become more significant than Internet price when we explain a group of technologies using an index. There is also a strong positive and significant relationship between regulatory quality and the digitalization index for highly digitalized countries (Hypothesis 6).

For countries showing high levels of digitalization, factors such as high regulation quality, higher infrastructure and economic development have a positive influence on the digitalization levels, while the percentage of the population between 15 and 64 years is negatively related. For lowly digitalized countries, nevertheless, the main factor is infrastructure, followed by GDP. For middle-digitalization economies, the main factor is the GDP followed by the population between 15 and 64 years and education.

In a second step we aimed to capture the multidimensional character of the digital divide. For this reason we study different combinations of technologies for each group of countries and explain them using CCA and a variety of variables. We test the hypotheses that different patterns of digitalization associated with different groups of countries can be detected (Hypothesis 7) and explained by economic development (Hypothesis 8) and demographic, social and institutional variables (Hypothesis 9). All the hypotheses have been supported. However, the results depend on the group of countries analyzed.

We have found two different patterns both for highly digitalized countries. A first pattern is closely related to the general use of the Internet, while the second is associated with the development of e-commerce. The general use of the Internet is mainly related to infrastructure and population between 15 and 64 years, and to a lesser extent by educational level and Internet prices. The development of e-commerce is positively related to economic development, regulatory quality and Internet prices. This means that in countries with high development of e-commerce there is also a high level of economic development, regulatory quality and Internet prices.

For middle-digitalization countries the combination of technologies is not so clearly associated with a specific pattern. The combination of PC, Internet use and mobile phone users shows a less defined pattern compared with developed countries. This pattern is positively associated with economic development and infrastructure.

According to these results, some policy strategies might boost digital diffusion depending on the level of ICT adoption. Both supply and demand-side initiatives should encourage digital diffusion. For highly digitalized countries, the promotion of wide Internet use should be based on the adoption of different measures related to infrastructure, prices and educational levels. From a supply perspective, different types of actions may improve digital infrastructure and reduce prices. These include deregulation and competition in telecommunications infrastructure and services, the development of public–private partnerships to develop new infrastructure at regional levels, and the development of attractive prices through local subsidies and flat-fee subscription models. From a demand perspective, measures to encourage education would have a very positive impact on Internet diffusion. It could be relevant to promote digital literacy in order to boost a more productive use of the Internet, in particular.

The promotion of e-commerce in highly digitalized countries may be encouraged through additional policy actions. First, any measure that facilitates the building of a secure environment for transaction in online markets would have a positive impact on e-commerce diffusion. Additionally, efficient and favorable business conditions among telecommunications operators should boost them to provide new and innovative digital services and applications. New digital services and contents in the e-government, e-health, e-learning and e-business fields should be more effective than any subsidy to promote e-commerce diffusion.

For middle-digitalization countries, specific policy actions could foster the use of PC, the Internet and mobile phones. First, economic growth in middle-digitalization countries will have a powerful influence on ICT diffusion. Second, the improvement of telecommunication infrastructures is another key point to boost digital diffusion. Due to the traditional lack of financial resources in this type of countries, competition and private–public partnerships should be encouraged. Competitive measures could focus on service competition more than on infrastructure competition, following Höffler (2007). Nevertheless, pro-competitive policies might not be enough to encourage infrastructure investment. The development of regional initiatives and complementary public efforts to guarantee an equitable access to ICT in urban and rural areas should be also highlighted. As Turk et al. (2008) have emphasized, regional initiatives appear to have been more successful than country-wide strategic plans in these countries. For example, regional initiatives in rural areas to develop telecommunications infrastructure in libraries, schools and community centers might be an inexpensive way to boost digital diffusion in many middle-digitalization countries.

The lack of data for many of the explanatory variables is one of the main limitations of this study. It limits the possibility of dynamic analysis to investigate how the evolution of the different economic, institutional, social and demographic factors affects the evolution of the digitalization index.

Given the fact that it requires the inclusion of a large number of countries, canonical correlation analysis did not allow us to determine differences in digitalization levels among less developed countries. Further research should attempt to discriminate according to economic development levels. Apart from analyzing ICT adoption and access, the differences in ICT use should be studied, accounting for the necessary acquisition of skills needed to promote a more productive use of some technologies, such as the Internet.

Appendix

See Tables 10, 11 and 12.

Table 10 Selection of studies on the digital divide

Author	Study setting/methodology	Dependent/independent variables	Findings
Hargittai (1999)	Period: 1998 Sample: 18 countries—OECD OLS regression	*Internet hosts/10.000* GDP, Gini coefficient, gross enrollment ratio, English proficiency, the average cost of a 20 h monthly Internet access basket, phone density	Economic wealth and telecommunications, predictors of Internet connectivity. Lack of competence in telecommunication sector has negative impact. Price of access and English proficiency are not significant
Dasgupta et al. (2001)	Period: 2000 Sample: 44 countries from Africa, Asia and Latin America Gompertz analysis—regression statistics	*Internet subscribers/telephone lines* Urban population, income p/c, index of competition policy, regional dummy variables	There is no gap in Internet intensity between high- and low-income countries. Influence of policy differences
Kiiski and Pohjola (2002) (1)	Period: 1995–2000 Sample: 23 countries—OECD Gompertz model	*Internet hosts per capita* Speed of diffusion, GDP p/c, Internet access cost, telecom competition, years of schooling, English proficiency, Nordic country dummy, southern country dummy, dummy for Mexico and Turkey (low income)	GDP p/c and Internet cost explain Internet connectivity. The regulation in Telecom is relevant only if it lowers the access cost. Investment in education is not significant
Kiiski and Pohjola (2002) (2)	Years: 1995–2000	Equation 1. *Internet hosts* p/c Speed of diffusion, GDP p/c, telephone access cost, years of schooling	GDP per capita and the average years of schooling are statistically significant. Investment in university education is more relevant than years of schooling
	Sample: developed and developing countries	Equation 2. *Infrastructure* PC/1.000, telephone mainlines/100	
	Three-stage least squares	Equation 3. *Price* Share of telecom investment in GDP. Share of telecom revenue in GDP. GDP per capita	
Quibria et al. (2003)	Period: 2000 Sample: 157 developed and developing countries	*Mobile phone, fax machine, TV, telephone mainline, Internet use, PC*	PC and Internet use correlated with Income but not with population size. Cell phones, TV, telephone mainlines adoption appears to increase with the rise of population
	Regression analysis	GDP per capita, size population	

Table 10 continued

Author	Study setting/methodology	Dependent/independent variables	Findings
Wong (2002)	11 Asian countries—1998/Asian as whole—1994–1998/32 non Asian countries OLS regression analysis	Eight indicators of ICT diffusion: PCs/1.000, computing power in millions of instruction per second (MIPS)/ 1.000, Internet hosts/1.000, secure e-commerce hosts/ 1.000, Telephone lines/1.000, Mobile phone/1.000, estimated electronics goods consumption p/c, estimated ICT expenditure p/c	Higher disparities in ICT diffusion in Asian countries than in non-Asian ones. Weaker correlation between competitiveness in electronic production and ICT, while higher correlation between GDP p/c and ICT diffusion
Beilock and Dimitrova (2003)	Period: 2000 Sample: 105 countries—developed and developing countries Tobit analysis and regression analysis	*Internet users/10.000* GNP (p/c income), telephones/1.000, PC/1.000, freedom house's index	GNP is the main determinant of Internet use
Tanner (2003)	Period: 1990–2001 Sample: 19 countries—Latin America OLS regression and FGLS	*Internet users/1.000* Gross national income p/c, Gini coefficient, telephone lines/100 people, privatization of state-owned telephone companies/liberalization of the telecommunications market, policies to reducing dial-up charges, the establishment of public access centers or telecenters, the year the country connected to the Internet Control variables: Internet users (1-year lag), OECD Internet users	There is a gap within Latin America over Internet access. Internet use is associated with a country's wealth and the development of telecommunications infrastructure. The government policy with the strongest influence on increasing access is changing the tariff structure. Market liberalization and the worldwide spread of the Internet are also associated with increased access
Dewan et al. (2005)	Period: 1985–2001 Sample: 40 countries—22 developed and 18 developing countries OLS and quantile regressions	*Mainframes, PCs and Internet* Telephone lines/1.000, telephone subscription cost/month, cost of local call, size of urban population (% total population), GDP p/c (PPA), years of schooling, trade (% GDP)	National income explains IT penetration. There are significant differences in the effects of other demographic and economic factors across countries at different stages of IT adoption. Factors that previously may have been expanding the divide with earlier technologies are narrowing the gap. The dynamics of the diffusion of the Internet appear to be different from the diffusion of earlier technologies

Table 10 continued

Author	Study setting/methodology	Dependent/independent variables	Findings
Guillén and Suárez (2006)	Years: 1997–2001	*Internet users/100*	Regulatory, political and sociological variables predict digital divide
	Sample: 118 countries (developed and developing countries)	Core/semi-periphery countries, competition, democracy, expenditure on tourism abroad by residents	
	Linear regression with panel-corrected standard errors	Control variables: GDP p/c, telephone lines/100, the cost of internet access, literacy	
	Regression equations using heteroskedasticity-consistent standard errors. Generalized least squares		
Oyelaran-Oyeyinka and Lal (2005)	Period: 1995–2000	Equation 1—D.V.: Internet use. I.V.: density of internet hosts, density of personal computers	Telecommunications infrastructure is vital for Internet diffusion. Impact of educational qualifications is twice that of income. Gaps in education and income lead to access inequality
	Sample: 40 countries—sub-Saharan Africa	Equation 2—D.V.: Density of Internet hosts. I.V.: telephone density, investment in telecommunication infrastructure, human capital	
	Univariate analysis: maximum likelihood method	Equation 3—D.V.: Telephone density. I.V.: internet use, GDP	
Andonova (2006)	Period: 1985–2001	*Telephone lines p/c, PCs with Internet access p/c, Internet hosts p/c, mobiles p/c*	Internet access depends more on institutional environment than mobile phones. Human capital and the degree of urbanization determinate the diffusion of information technology
	Least squares hierarchical regression	POLCON, political rights, civil liberties (C.V.), Business charge, business subscription, cellular charge, cellular phone subscribers, cellular subscription, cost of a 3 min call, GDP p/c, illiteracy, Internet hosts, Internet users, main lines, urban population, residential charge, and residential subscription	

Table 10 continued

Author	Study setting/methodology	Dependent/independent variables	Findings
Chinn and Fairlie (2007)	Period: 1999–2001 Sample: 161 countries (developed and developing countries) Panel data regression model	*PC/100—Internet users/100* Main telephone lines/100p., monthly telephone subscription charge, cost of a 3-min call, Electric power consumption (kwh p/c), Population (aged 0–14, 65–), urban population, gross national income p/c, years of schooling, regulatory quality, trade, illiteracy rate, average computer penetration rate	Per capita income explains the gap in computer and Internet use. Telecom infrastructure, quality of regulation, also important. Education is relevant for PC but not for Internet use
Corrocher and Ordanini (2002)	Period: 2000 and 2001 Sample: 10 developed countries Factor analysis	Factor index measuring six dimensions	There is a digital divide between developed countries. Different patterns of digitalization by country
Bagchi (2005)	Period: 1995 and 2001 Sample: OECD (30 countries) and ECLAC (33 countries) Principal component analysis (PCA), ANOVA and OLS	*Telephone/1.000, cell-phone/1.000, PC/1.000 and Internet usage/1.000 and the digital distance of a nation from the US* GDP p/c, inflation, information technology expenditure, Gini index, secondary education average, illiteracy level, trust index, urbanization, ethno-linguistic fractionalization index, television/1,000	Results are different for nations belonging to different clusters. Many indicators impact the digital divide index in different and significant ways across nations and over time

Table 11 Digitalization index over log transformed variables: factor analysis results

Factor	Eigen-value	Total variance explained		Variables	Factor 1	
		Percent of variance	Cumulative percent of variance		Factor loadings	Communality
1	5.296	88.262	88.262	PC	0.946	0.894
2	0.217	3.622	91.884	IIB	0.961	0.924
3	0.198	3.303	95.186	SIS	0.946	0.895
4	0.133	2.212	97.398	IU	0.955	0.912
5	0.085	1.419	98.817	BBS	0.913	0.834
6	0.071	1.183	100	MPS	0.915	0.837
KMO measure of sampling adequacy			0.923			
Barlett test of sphericity			991.2 [$p = 0.000$]			

Extraction method is principal component analysis

Table 12 Digitalization index by country (log transformed variables)

High-digital. group	Index	Middle-digital. group	Index	Low-digital. group	Index
Sweden	1.434	Costa Rica	0.516	Egypt, Arab Rep.	−0.465
Denmark	1.432	Uruguay	0.396	Ukraine	−0.507
Netherlands	1.402	Argentina	0.392	Sri Lanka	−0.509
Switzerland	1.368	Romania	0.383	Senegal	−0.517
United Kingdom	1.347	Brazil	0.374	Azerbaijan	−0.518
United States	1.321	Mauritius	0.359	Namibia	−0.522
Hong Kong, China	1.321	Turkey	0.311	Nicaragua	−0.527
Canada	1.308	Trinidad and Tobago	0.308	Indonesia	−0.589
Singapore	1.305	Mexico	0.306	Zimbabwe	−0.593
Norway	1.293	Panama	0.290	Botswana	−0.606
Austria	1.267	Lebanon	0.239	Vietnam	−0.641
Finland	1.260	Russian Federation	0.221	Kyrgyz Republic	−0.697
Germany	1.217	Peru	0.169	Algeria	−0.707
Australia	1.204	Venezuela, RB	0.130	Honduras	−0.749
Belgium	1.182	South Africa	0.127	Albania	−0.778
Ireland	1.178	Saudi Arabia	0.122	India	−0.779
New Zealand	1.152	Thailand	0.081	Swaziland	−0.823
Japan	1.149	Jordan	0.070	Pakistan	−0.957
Estonia	1.149	Bosnia and Herzegovina	0.042	Cote d'Ivoire	−1.018
Israel	1.146	Colombia	0.021	Togo	−1.026
France	1.085	El Salvador	−0.011	Uzbekistan	−1.086
Italy	1.025	Oman	−0.023	Zambia	−1.104
Korea, Rep.	1.025	Philippines	−0.105	Haití	−1.107
Slovenia	1.022	Gabon	−0.105	Kenya	−1.126

Table 12 continued

High-digital. group	Index	Middle-digital. group	Index	Low-digital. group	Index
Spain	1.016	Tunisia	−0.110	Cameroon	−1.168
Czech Republic	0.877	Ecuador	−0.118	Papua New Guinea	−1.317
Portugal	0.826	China	−0.133	Nigeria	−1.357
Latvia	0.806	Moldova	−0.159	Cuba	−1.362
Slovak Republic	0.805	Morocco	−0.170	Cambodia	−1.385
United Arab Emirates	0.800	Mongolia	−0.175	Burkina Faso	−1.463
Hungary	0.763	Dominican Republic	−0.187	Uganda	−1.478
Chile	0.670	Paraguay	−0.217	Tanzania	−1.532
Poland	0.653	Bolivia	−0.271	Mozambique	−1.548
Croatia	0.651	Georgia	−0.352	Madagascar	−1.624
Lithuania	0.630	Armenia	−0.389	Burundi	−1.686
Jamaica	0.585	Macedonia, FYR	−0.414	Nepal	−1.735
Kuwait	0.583	Iran, Islamic Rep.	−0.433	Malawi	−1.759
Malaysia	0.572	Guatemala	−0.437	Bangladesh	−1.801
Greece	0.566			Ethiopia	−2.277

References

Andonova, V. (2006). Mobile phones, the Internet and the institutional environment. *Telecommunications Policy, 30*(1), 29–45.

Andonova, V., & Diaz-Serrano, L. (2009). Political institutions and telecommunications. *Journal of Development Economics, 89*(1), 77–83.

Bagchi, K. (2005). Factors contributing to global digital divide: Some empirical results. *Journal of Global Information Technology Management, 8*(1), 47–65.

Beilock, R., & Dimitrova, D. V. (2003). An exploratory model of inter-country Internet diffusion. *Telecommunications Policy, 27*(3–4), 237–252.

Chinn, M. D., & Fairlie, R. W. (2007). The determinants of the global digital divide: A cross-country analysis of computer and internet penetration. *Oxford Economic Papers, 59*(1), 16–44.

Corrocher, N., & Ordanini, A. (2002). Measuring the digital divide: A framework for the analysis of cross-country differences. *Journal of Information Technology, 17*(1), 9–19.

Crenshaw, E. M., & Robison, K. K. (2006). Globalization and the digital divide: The roles of structural conduciveness and global connection in Internet diffusion. *Social Science Quarterly, 87*(1), 190–207.

Dasgupta, S., Lall, S., & Wheeler, D. (2001). *Policy reform, economic growth and the digital divide. Development research group. An econometric analysis*. (Policy research working paper series 2567). Washington, DC: World Bank.

Dasgupta, S., Lall, S., & Wheeler, D. (2005). Policy reform, economic growth and the digital divide. *Oxford Development Studies, 33*(2), 229–243.

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.

Dewan, S., Ganley, D., & Kraemer, K. L. (2005). Across the digital divide: A cross-country multi-technology analysis of the determinants of IT penetration. *Journal of the AIS, 6*(4), 409–432.

Donner, J. (2008). Research approaches to mobile use in the developing world: A review of the literature. *The Information Society, 24*(3), 140–159.

Fujita, M., & Mori, T. (2005). Frontiers of the new economic geography. *Papers in Regional Science, 84*(3), 377–405.

Guillén, M. F., & Suárez, S. L. (2006). Explaining the global digital divide: Economic, political and sociological drivers of cross-national Internet use. *Social Forces, 84*(2), 681–708.

Hair, J. F. Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). New Jersey: Prentice-Hall.

Hargittai, E. (1999). Weaving the Western Web: Explaining differences in Internet connectivity among OECD countries. *Telecommunications Policy, 23*(10–11), 701–718.

Höffler, F. (2007). Cost and benefits from infrastructure competition. Estimating welfare effects from broadband access competition. *Telecommunications Policy, 31*(6–7), 401–418.

International Telecommunication Union (ITU) (2006). *World telecommunication development report and database 2006: Measuring ICT for social and economic development*. Geneva: ITU.

International Telecommunication Union (ITU) (2008). *2008 Report on the world summit on the information society stocktaking*. Geneva: ITU.

James, J. (2007). From origins to implications: Key aspects in the debate over the digital divide. *Journal of Information Technology, 22*(3), 284–295.

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Upper Saddle River, NJ.: Prentice-Hall.

Karshenas, M., & Stoneman, P. (1995). Technological diffusion. In P. Stoneman (Ed.), *Handbook of the economics of innovation and technological change* (pp. 265–297). Oxford: Blackwell.

Kiiski, S., & Pohjola, M. (2002). Cross-country diffusion of the Internet. *Information Economics and Policy, 14*(2), 297–310.

Lee, S., & Brown, J. (2008). Examining broadband adoption factors: An empirical analysis between countries. *Info, 10*(1), 25–39.

Liu, M., & San, G. (2006). Social learning and digital divides: A case study of Internet technology diffusion. *Kyklos, 59*(2), 307–321.

Mansfield, E. (1968). *Industrial research and technological innovation*. New York: Norton.

OECD (2001). *Understanding digital divide*. Paris: OECD.

OECD (2005). *Working party on indicators for the information society. DSTI/ICCP/IIS(2005)4*. Paris: OECD.

Oyelaran-Oyeyinka, B., & Lal, K. (2005). Internet diffusion in sub-Saharan Africa: A cross-country analysis. *Telecommunications Policy, 29*(7), 507–527.

Pick, J., & Azari, R. (2008). Global digital divide: Influence of socioeconomic, governmental, and accessibility factors on information technology. *Information Technology for Development, 14*(2), 91–115.

Quibria, M. G., Shamsun, A. N., Tschanh, T., & Reyes-Macasaquit, M. (2003). Digital divide: Determinants and policies with special reference to Asia. *Journal of Asian Economics, 13*(6), 811–825.

Rogers, E. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.

Sciadas, G. (2005). Info states across countries and over time: Conceptualization, modeling, and measurements of the digital divide. *Information Technology for Development, 11*(3), 299–304.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins.

Tanner, E. (2003). Bridging Latin America's digital divide: Government policies and Internet access. *Journalism Mass Communication Quarterly, 80*(3), 646–665.

Turk, T., Blazic, B., & Trkman, P. (2008). Factors and sustainable strategies fostering the adoption of broadband communications in an enlarged European Union. *Technological Forecasting and Social Change, 75*(7), 933–951.

Vehovar, V., Sicherl, P., Hüsing, T., & Dolnicar, V. (2006). Methodological challenges of digital divide measurement. *The Information Society, 22*(5), 279–290.

Vicente, M. R., & López, A. J. (2006). A multivariate framework for the analysis of the digital divide: Evidence for the European Union-15. *Information and Management, 43*(6), 756–766.

Wallsten, S. (2005). Regulation and Internet use in developing countries. *Economic Development and Cultural Change, 53*(1), 501–524.

Wong, P. K. (2002). ICT production and diffusion in Asia: Digital dividends or digital divide? *Information Economics and Policy, 14*(2), 167–187.

World Bank (2006). *Information technologies and development*. Geneva: World Bank.