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

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 and low-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.

Classification code (JEL): L86, L96, O33.

Margarita Billón
Universidad Autónoma de Madrid
Departamento de Estructura Económica y Economía del Desarrollo
Carretera de Colmenar Viejo, Km 15
28049 Madrid
Tfno.: +34 91 497 8777
Fax +34 91 497 1971
E-mail: margarita.billon@uam.es

Fernando Lera-López
Universidad Pública de Navarra
Departamento de Economía
Campus Arrosadía s/n
31006 Pamplona
Tfno. 948-169351
Fax: 948-169721
E-mail: lera@unavarra.es

Rocío Marco
Universidad Autónoma de Madrid
Departamento de Economía Aplicada-Estadística
Carretera de Colmenar Viejo, Km 15
28049 Madrid
Tfno.: +34 91 497 2899
Fax: +34 91 497 4076
E-mail: rocio.marco@uam.es

1. Introduction

Given the fact that information and communication technologies (ICTs) 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.

ICT 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) 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 seeks to identify and explain the differences in the digitalization levels between different groups of countries as well as the relationships between specific combinations of technologies and the factors explaining them .

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 or explain the digital divide (Corrocher and Ordanini, 2002; Bagchi, 2005) in that the present study includes different types of technologies capturing ICT use and infrastructure, 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 UE-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 (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 Section 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

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, Figure 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.

We show that mobile phones (MPS), the Internet (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.

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 Figure 1, indicators associated with higher levels of infrastructure such as International Internet Bandwidth (IIB) and the number of Secure Internet Servers (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

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).

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 (Oyelaran-Oyeyinka and Lal, 2005; Dasgupta et al., 2005), while others include a combination of developing as well as developed countries (Kiiski and Pohjola, 2002; Dewan et al., 2005; Guillén and Suárez, 2006; Bagchi, 2005; Chinn and Fairlie, 2007; Andonova and Diaz-Serrano, 2008; Pick and Azari, 2008).

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 shown in Figure 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. 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 five-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; Dewan et al., 2005). According to the diffusion theories (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 about new technologies. Demographic variables such as population size, population 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 (Hargittai, 1999; Guillén and Suárez, 2006). Andonova and Diaz-Serrano (2008), 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. 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.

4. Research model and methodology

Our research model seeks to determine the factors explaining different levels of digitalization. We create an index summarizing variables related to the access and use of several information and communication technologies to capture the level of digitalization. Following the empirical evidence available, we study whether the digitalization level of different groups of countries is explained by economic, socio-demographic and institutional factors.

To capture the level of digitalization, we begin by using a factorial analysis for the digitalization variables selected. 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 that purpose, given the multidimensional nature of the digital divide and the variety of factors affecting it, we use canonical correlation analysis (CCA). It provides an additional contribution with respect to that of the multiple regression analysis, commonly used in other studies. 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.

CCA seeks to identify and quantify the association between two groups of variables (Johnson and Wichern, 2002). 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 *x** and *y** provide simple summary measures of the set of explanatory variables *X* and the digitalization variables *Y*. Set:

$$x^* = \boldsymbol{X\alpha} = \sum_{i=1}^{p} \alpha_i x_i \qquad\qquad y^* = \boldsymbol{Y\beta} = \sum_{j=1}^{q} \beta_j y_j$$

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:

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

$\boldsymbol{V}_{11}$= *Cov* (***X***), $\boldsymbol{V}_{22}$= *Cov* (***Y***) and $\boldsymbol{V}_{12}$= *Cov* (***X,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 equation [1]. The second pair of canonical variables is the pair of linear combinations x_2^*, y_2^* with unit variances that maximize correlation [1] 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.

5. Data and variables

The final database includes 142 countries for the year 2004[1]. 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.

The set of dependent variables includes various types of telecommunications technologies. Some ICT, such as personal computers (PCs), international Internet bandwidth (IBB) and secure Internet servers (SIS) are related

[1] 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.

to the infrastructure needed to support the use of other technologies, while others are indicators of ICT use: Internet users (IU), Broadband subscribers (BBS) and mobile phone subscribers (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). International Internet bandwidth refers to the 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 (see Lee and Brown, 2008 for a recent review of the literature). 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; Guillén and Suárez, 2006; Oyelaran-Oyeyinka and Lal, 2005; Bagchi, 2005; 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.

The set of explanatory variable includes 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 (Hargittai, 1999; Kiiski and Pohjola, 2002; Quibria et al., 2003; Dewan et al., 2005; Guillén and Suárez, 2006; Bagchi, 2005). 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 telephone mainlines lines per 100 people (TM) has been incorporated as an explanatory variable. We have employed school life expectancy (SLE) to measure the impact of education and the percentage of population between 15-64 years of age (POP2) to reflect the influence of age distribution on ICT adoption. We expect that education has a positive influence on ICT diffusion (Kiiski and Pohjola, 2002; Quibria et al., 2003; Crenshaw and Robison, 2006). As a higher percentage of elderly has been shown to have a negative effect on ICT adoption (Chinn and Fairlie, 2007), we expect a positive relationship between the population between 15 and 64 years and ICT diffusion. Regarding institutional factors, we have considered the cost of Internet use (IP20, Internet price for 20 hours of use) and the quality of regulation (regulatory quality, 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, although this would contrast the results of Chinn and Fairlie (2007).

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 Governance Indicators Database (World Bank) (see Table 2).

The exploratory analysis obtained from the main descriptive statistics evidences 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[2]. 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: broadband subscribers (BBS), Internet users (IU), secure Internet servers (SIS), personal computers (PCs), international Internet bandwidth (IIB) and mobile phone subscribers (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[3]. 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 3). Our factor extraction relied on principal component analysis. By eigenvalue criterion, we finally only consider one factor

[2] 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.

[3] 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.

(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 ($r = 0.914$). Mean values of GDP and the digitalization index divide 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.

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. The results of the factor analysis are shown in Appendix I. After extracting principal components, we obtain one relevant factor which explains 88.26% of the total variance.

We divide the sample of countries into thirds, according to the digitalization index score obtained. The three groups are shown in Appendix II. 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 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 percentage of the population between 15-64 years of age (POP2), school life expectancy (years) (SLE) representing education level, the number of telephone mainlines lines (TM) to capture the role of infrastructure and Internet price (IP20) and regulatory quality (RQ) as institutional variables.

Table 4 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 4 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. 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 15-64 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 it 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.

Our dependent variables are those included in the digitalization index (PC, IU, IIB, SIS, 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 5 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 twenty hours of Internet use (IP20) is negatively associated with all the digitalization variables, showing the weakest correlation with secure Internet servers (SIS) (- 0.195).

The CCA results in Table 6 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 the Chi-square sequential test statistic[4]. 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 (2002) 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.

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.

[4] The null and alternative hypotheses for assessing the statistical significance of the first *k*th canonical correlations are

$$H_0^k : \rho_1 \neq 0, \rho_2 \neq 0, ..., \rho_k \neq 0, \rho_{k+1} = 0, ..., \rho_p = 0$$

$$H_1^k : \rho_i \neq 0, \quad \text{for same } i \geq k+1$$

which has an approximate Chi-square distribution assuming multivariate normal data.

Table 7 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 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 (telephone mainlines, TM), to a lesser extent by education (SLE) and GDP, and negatively by Internet prices (IP20). These results are in line with those of other, previously mentioned studies where Internet use, for example, is explained by demographic factors, some socioeconomic variables such as education, and Internet prices (Kiiski and Pohjola, 2002; Corrocher and Ordanini, 2002; Dewan et al., 2005; Guillén and Suárez, 2006; Oyelaran-Oyeyinka and Lal, 2005; Andonova, 2006). The dummies included indicate that this dimension refers to highly digitalized countries.

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 Electronic Commerce and its correlation with IP20 is negative although weak (see Table 5). 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. Along with IP20, the second dimension is explained by GDP and, to a lesser extent, by RQ.

As with the first dimension, the second pattern is also related to highly-digitalized countries, but the negative coefficient points to countries with a relative lower digitalization level. 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, Slovak Republic and Estonia. Countries included in the second digital diffusion pattern are, for example, United Kingdom and Switzerland.

The relationships between the two set 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.

8. Discussion and conclusions

We have found that the variables explaining the level of digitalization differ according to the group of countries considered. 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 15-64 years and education.

The only factor that seems to have a significant effect on digital development for all country groups is economic development. Nevertheless, its relevance varies according to the group. The results show that its role is more significant for middle-digitalization countries and less relevant for highly-digitalized countries where other non-economic variable become more important.

To capture the multidimensional character of the digital divide, we have studied different combinations of technologies for each group of countries and explain them using CCA and a variety of variables. Two different patterns emerge for highly-digitalized countries. A first pattern is closely related to the general use of the Internet, while a second is associated with the development of e-commerce. Different patterns of digitalization are also explained by different factors. The general use of the Internet is mainly explained by infrastructure and the population between 15-64 years, and to a lesser extent by educational level and Internet prices. The development of e-commerce is positively explained by 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.

The results show that 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 countries other than high-digitalized ones, 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 middle and less developed countries. Further research should attempt to discriminate according to economic development levels. We might to carry out the CCA only for the middle- and lowly-digitalized economies.

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.

References

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

Andonova, V., Diaz-Serrano, L. (2008). Political institutions and telecommunications. *Journal of Development Economics*, doi:10.1016/j.jdeveco.2008.09.006.

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, 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.

Johnson, R.A., Wichern D.W. (2002). Applied multivariate statistical analysis. (5th ed). 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.

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. (5^{th} 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.

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.

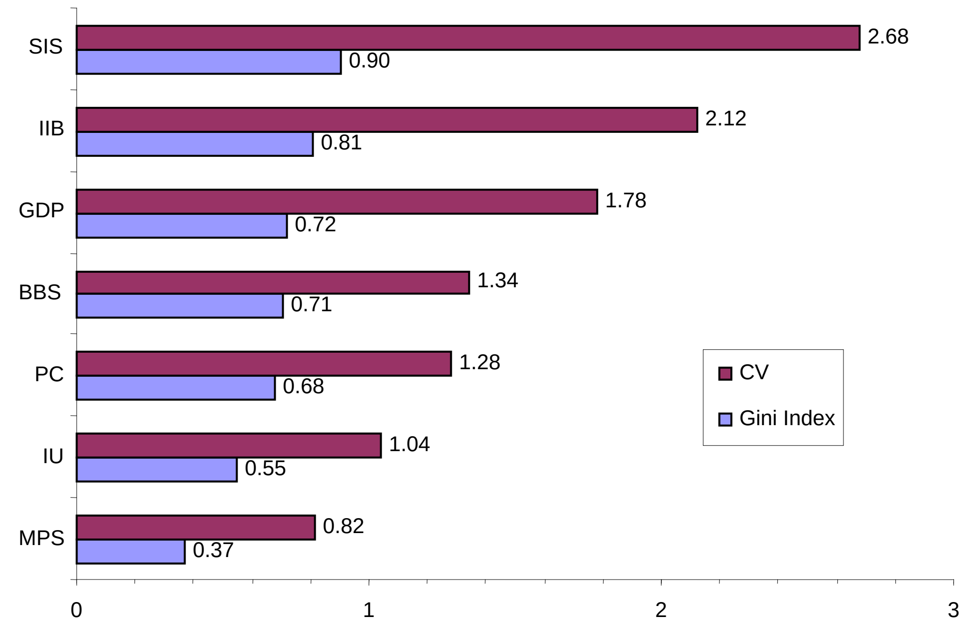
Vehovar, V., Sicherl, P., Hüsing, T., Dolnicar, V. (2006). Methodological challenges of digital divide measurement. *The Information Society*, 22(5), 279-290.

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.

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.

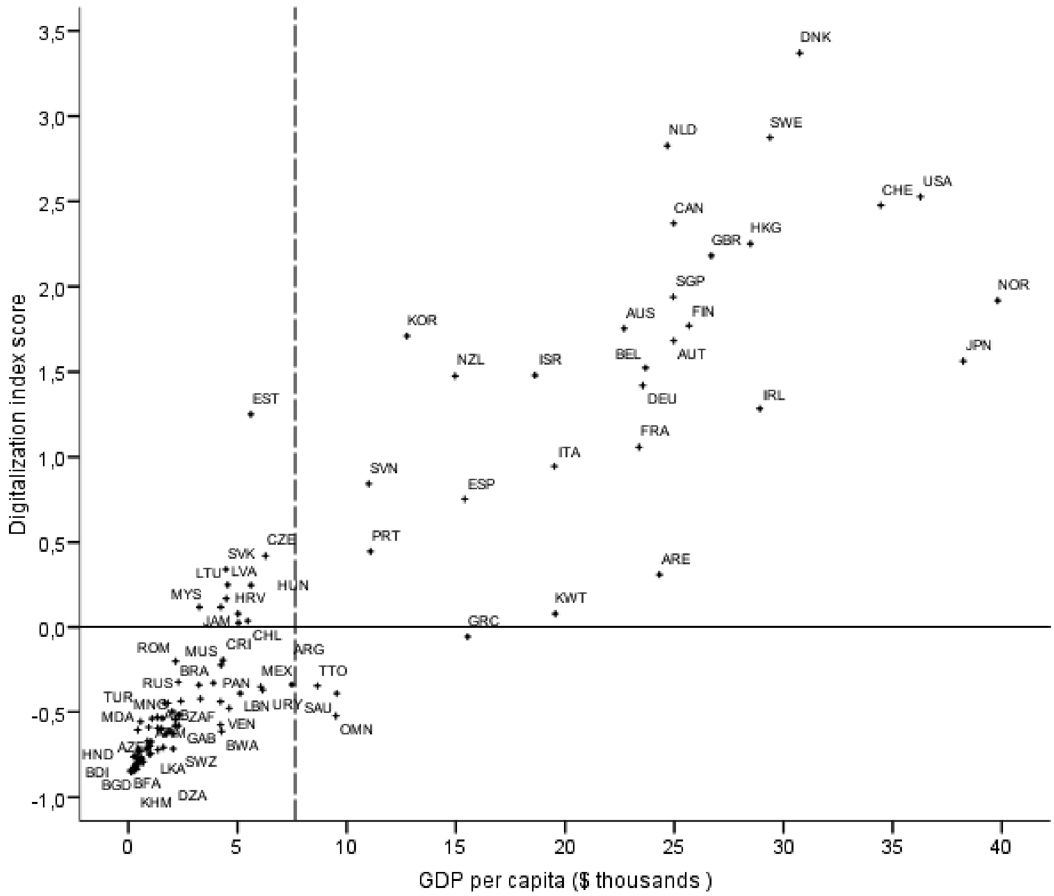
Figure 1: *Dispersion and inequalities between countries in ICT adoption*



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)

Figure 2: *GDP per capita vs Digitalization Index*



Source: Author´s work

Table 2: *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.
IP20	142	3.850	167.510	37.111	27.870	30.731	Internet price, 20 hours 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 work.

Table 3: *Digitalization index: Factor analysis results*

Factor	Eigenvalue	Total variance explained: Percent of variance	Cumulative percent of variance
1	4.667	77.79	77.79
2	0.542	9.03	86.82
3	0.447	7.45	94.26
4	0.213	3.55	97.81
5	0.087	1.45	99.26
6	0.044	0.74	100
KMO measure of sampling adequacy			0.815
Barlett test of sphericity			804.1 [p=0.000]

Variables	Factor 1: Factor Loadings	Communality
BBS	0.918	0.843
IU	0.941	0.886
MPS	0.838	0.702
PC	0.973	0.947
IIB	0.770	0.594
SIS	0.834	0.696

Note: Extraction method by principal component analysis

Table 4: *Regression result for the Digitalization Index*

	High digital		**Middle digital**		**Low digital**	
	β	Stand. β	β	Stand. β	β	Stand. β
IP20	-0.025 (0.045)	-.044	0.106 (0.070)	0.207	0.097 (0.083)	0.152
GDP	0.104** (0.046)	0.284	0.185*** (0.050)	0.515	0.230** (0.091)	0.395
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
TM	0.223** (0.101)	0.301	0.020 (0.073)	0.040	0.238*** (0.071)	0.649
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)*

Table 5: *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(**)	,876(**)	,874(**)	,911(**)	,910(**)	,915(**)
SLE	0.706(**)	,804(**)	,784(**)	,833(**)	,770(**)	,753(**)
POP2	0.744(**)	,788(**)	,734(**)	,775(**)	,738(**)	,608(**)
TM	0.799(**)	,904(**)	,828(**)	,894(**)	,847(**)	,818(**)
RQ	0.669(**)	,677(**)	,749(**)	,714(**)	,775(**)	,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).*

Table 6: *Canonical correlations*

Measures of overall model fit				
Multivariate Statistics		*Value*	*F value*	*p-value*
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 Correlation Test				
Canonical pair	*Canonical correlation*	*Chi-Sq Test*	*df*	*p-value*
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 Redundancy Analysis				
Canonical pair		*Squared Correlation*	*Variance extracted in set Y*	*Redundancy Measure* y^*_i/x^*_i
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

Table 7: *Canonical variables*

	y^*_1		**y^*_2**	
Set Y	*Canonical coefficients*	*Canonical loadings*	*Canonical coefficients*	*Canonical loadings*
BBS	0.307	0.860	-0.423	0.235
IU	0.305	0.906	-0.720	0.231
SIS	-0.846	0.741	2.233	0.626
PC	0.737	0.902	-0.884	0.288
IIB	0.339	0.867	-0.168	0.382
MPS	0.120	0.817	0.374	0.401
	x^*_1		**x^*_2**	
Set X	*Canonical coefficients*	*Canonical loadings*	*Canonical coefficients*	*Canonical loadings*
IP20	-0.446	-0.657	0.924	0.507
GDP	0.005	0.843	1.035	0.458
SLE	0.230	0.840	-0.316	0.146
POP2	0.376	0.876	-0.648	0.000
TM	-0.325	0.868	0.989	0.271
RQ	-0.032	0.655	0.304	0.478
HIGHDIG	0.750	0.708	-0.689	0.377
MIDDDIG	0.309	0.041	-0.155	-0.058

Note: Variables are log transformed. Canonical coefficients are standardized

Appendix I: *Digitalization index over log transformed variables: factor analysis results*

Factor	Eigenvalue	Total variance explained: Percent of variance	Cumulative percent of variance
1	5.296	88.262	88.262
2	0.217	3.622	91.884
3	0.198	3.303	95.186
4	0.133	2.212	97.398
5	0.085	1.419	98.817
6	0.071	1.183	100
KMO measure of sampling adequacy			0.923
Barlett test of sphericity			991.2 [p=0.000]

Variables	Factor 1: Factor Loadings	Communality
BBS	0.913	0.834
IU	0.955	0.912
MPS	0.915	0.837
PC	0.946	0.894
IIB	0.961	0.924
SIS	0.946	0.895

Note: Extraction method is principal component analysis

Appendix II: *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
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