Faculté Jean Monnet
Université Paris Sud

International Conference - Conférence Internationale
"ICTs & Inequalities : the digital divides"
"TIC & Inégalités : les fractures numériques"
Paris, Carré des Sciences
18-19 novembre 2004

A multivariate approach to the digital divide in the European Union

María Rosalía Vicente Cuervo[1], Ana Jesús López Menéndez

University of Oviedo

Key words: digital divide, information and communication technologies (ICT), multivariate statistics, principal components analysis, cluster analysis

Abstract

Our paper attempts to response to the need of analysing the digital gap. With this aim we present a new approach to the measurement of the digital disparities between countries by means of multivariate statistical methods. On the basis of a set of ICT-related indicators in the European Union (EU), we use principal components and cluster analysis. Results lead to the identification of two factors and four groups of countries, reflecting the asymmetry in the development of the information society within the EU.

[1] Corresponding author: María Rosalía Vicente Cuervo, Departamento de Economía Aplicada, Campus del Cristo s/n, 33006, Oviedo – Asturias, Spain.
E-mail address: mrosalia@uniovi.es

1. Introduction

At the World Summit of the Information Society in December 2003, heads of state and government from all over the world declared that the global challenge for the new millennium was building a society where everyone could access and share information, enabling individuals, communities and peoples to achieve their full potential in promoting their development and improving their quality of life.

To achieve this goal, however, some obstacles need to be overcome, namely the extreme disparities of access both between nations and within nations. Before the suitable actions to implement for bridging those gaps can be decided, it is necessary to know about their magnitude.

Within this context the development of accurate indicators and measures of the digital disparities has become a matter of special importance over the last few years. This task however has not been free of difficulties due to different reasons, such as the absence of a clear definition of both the information society and the digital divide, the lack of a structured theoretical framework, and the shortage of adequate and harmonised data, among other reasons.

Our paper focus on the digital gap and attempts to identify its main components within the european context. With this aim we first try to clarify the notion of the digital divide (section 2). Then we explore patterns in levels of ICT access and use in the European Union (EU) (section 3). From this, we perform a multivariate analysis of the digital gap on data for 15 European countries (section 4). Finally we summarize our principal findings (section 5).

2. Defining the Digital Divide

As the diffusion of Information and Communication Technologies (ICTs) has taken place the issue of the digital divide has emerged to occupy a central position on both international and national forums.

It is a key matter to the extent that ICTs have the potential to foster economic growth and human development (UNDP, 2001). Moreover these technologies can help to achieve the Millennium Development Goals (UN ICT Task Force, 2003). Hence the digital divide represents the major threat to harness the opportunities offered by the ICTs.

According to the OECD (2001) "the term digital divide refers to the gap between individuals, households, businesses and geographic areas at different socio-economic levels with regard both to their opportunities to access information and communication technologies (ICTs) and to their use of the Internet for a wide variety of activities".

The notion of the digital divide is complex and multidimensional. As UN General Secretary Kofi Annan (2003) pointed out at the World Summit of the Information Society, "there are really several digital divides" that may overlap.

At first the issue of the digital divide was understood in binary terms: the gap between ICTs "haves" and "have-nots". But as the number of Internet users has grown, it has become quite relevant to look at differences on those who are online. Warschauer (2002) suggests that in addition to the physical sides of access, other factors such as content, language, literacy, education and institutional structures must also be taken into consideration when assessing the level of information and communication technology use in a community. Hargittai (2003) offers a refined understanding of the digital divide including five factors: quality of equipment, autonomy of use, presence of social support networks, experience and online skills.

The analysis of the digital disparities has focused on two main dimensions: the gap between countries (the international digital divide) and between groups within countries (the domestic digital divide).

The work done on the international digital divide shows that it is largely the consequence of the social and economic imbalances that exist between developing and industrial countries. Thus countries with lower income and lower educational attainment tend to show lower rates of ICT access and use when compared with higher income and higher education countries (Deutsche Bank, 2001; Hilbert, 2001a, 2001b; Vicente and López, 2004b).

Concerning the domestic digital divide some work has been done showing the risk of digital exclusion by women, the elder, those with lower income and educational attainment, with disabilities, living in rural areas, belonging to race or ethnical minorities (US Department of Commerce, 1995, 1998, 1999, 2000, 2002; OECD, 2001, 2002; Sciadas, 2002; Vicente and López, 2004a).

In this paper we focus on the first of these two dimensions, presenting a new approach to the analysis of the digital divide across countries. Our approach is consistent with the definition of the digital divide as multidimensional phenomena, which implies that making a deep analysis requires a broader spectrum of indicators than just Internet penetration rate.

Therefore in the next section we explore the digital disparities in the EU on the basis of a set of ICT-related indicators.

3. Overview of ICT diffusion in the European Union

The framework for ICT diffusion in the EU has been the *e*Europe Action Plan in its two versions: *e*Europe 2002 and *e*Europe 2005. This initiative is part of the Lisbon strategy designed to make the EU the most competitive and dynamic knowledge-based society by year 2010.

Within this context the development of accurate indicators of the progress towards the information society has become a matter of special importance. This task however has not been free of difficulties due to different reasons, such as the absence of a clear definition of both the information society and the digital divide, the lack of a structured theoretical framework, and the shortage of adequate and harmonised data, among other reasons.

In spite of the international efforts to develop statistical information about the digital society, there is usually a "trade-off" between breadth and depth in the selection of indicators. That is, the more indicators we try to use, the less number of countries we can include in our analysis.

Concerning the analysis of the digital divide done in this paper, we have used a set of ten indicators related to ICT infrastructure, e-commerce, access costs and ICT diffusion. The criteria used for their selection are: the availability of data for the EU-fifteen countries, the homogeneity of data and the quality and reliability of information sources. The required information is provided by Eurostat and reports to the year 2001.

The descriptive statistics shown in Table 1 reflect some important differences across countries. Thus the number of computers per 100 inhabitants goes from 8 in Greece to 56 in Sweden. Moreover the number of secure servers per million inhabitants goes from 17 in Greece to 155 in Luxemburg, that is, about 9 times larger.

Despite these differences, mean and median values are quite close. Besides, the uneven distribution of ICT does not impose any limitation to our analysis, as the methods employed do not make any distributional assumptions.

Table 1. Descriptive statistics

Code	Variables	Minimum	Maximum	Mean	Median	S.D.
PC	Number of computers per 100 inhabitants	8.00	56.00	33.73	36	14.07
TEL	Number of main telephone lines per 100 inhabitants	41.60	76.20	55.35	53.1	10.56
BRO	Number of broadband connections per 100 inhabitants	0.00	2.71	1.17	0.73	0.97
SER	Number of secure servers per million inhabitants	17.00	155.00	79.60	78	49.84
WEB	Percentage of businesses with a website	54.20	80.00	67.70	64.8	8.98
B2B	Percentage of businesses buying on line	5.30	41.90	24.95	28	10.85
COST	Internet dial up access costs for a residential user (off peak) (USD PPP)	16.97	38.98	28.95	28.19	5.41
HOH	Percentage of households connected to the Internet	11.70	64.30	40.17	43.6	14.91
PS	Percentage of public services on line	15.22	68.42	45.83	48.8	14.85
APOP	Percentage of active population using a computer for professional purposes	27.00	73.00	54.40	57	13.48

4. Methodology for the analysis of the Digital Divide

Over the last years significant efforts have been devoted to the understanding of the implications of ICT revolution for both social and economic spheres. Within this context the need for measures and analyses of the digital divide has emerged as a central issue. Thus several methodological approaches and statistical techniques have been used for this purpose.

In particular there have been an increasing number of attempts to capture the multiple dimensions of the information society by means of composite indicators[2]. Examples of such indices are the Information Society Index by IDC/ World Times (annual from 1995), the New Economy Index by the Progressive Policy Institute (PPI, 1998, 1999, 2001, 2002), the Technology Achievement Index developed by UNDP (2001), the Infostate Index by Orbicom (2002, 2003), and the Networked Readiness Index by the Center for International Development at Harvard University for the World Economic Forum (2002, 2003, 2004), among others.

The use of this kind of indicators has two main advantages: on the one hand they summarise complex and multidimensional phenomena such as the information society and the digital divide; on the other hand they are easier to interpret than many different indicators. Nevertheless they have also some limitations: they might give a simplistic picture of a country' situation and, moreover, rankings depend largely both on which indicators are included and which ones are excluded from the index.

Aware of these drawbacks UNESCO (2003) suggests that a more sensible approach would be to group countries of similar economic status together and to attempt analyses within these groups using the results coming from a multitude of well-chosen but different indicators that are not amalgamated into one entity.

In this context our analysis of the digital divide focuses on two main objectives. In the first place, we try to identify the dimensions that underlie a set of ICT-related indicators by means of principal components analysis[3]. In the second place, we group countries according to their digital development using cluster techniques.

---

[2] Composite indicators are mathematical combinations of a set of indicators.
[3] Correcher and Ordanini (2002) also used principal components analysis for measuring the digital divide across OECD countries. Contrary to our approach, they use this method as a procedure for assigning weights to a set of indicators in order to obtain a composite index.

This methodological approach represents one of the most valuable features of our work, since it captures the multiple dimensions of the digital divide, showing the interrelations that may exist between them.

Despite the fact that our area of analysis is the European Union, our approach could be transferred to any other set of countries or geographical areas.

4.1. Principal components analysis

The main purpose of principal components analysis is to reduce the complexity of the multivariate data into a small number of dimensions that explains most of the variation in the original variables.

Thus it allows the transformation of a given set of variables into a group of new components through linear combinations of the original variables. The extraction method ranks the new components according to the decreasing shares of explained variance.

This analysis involves various steps. In the first place, it is necessary to assess whether the variables are sufficiently correlated with each other so that applying this method can reduce them. In the second place, the number of factors that should be extracted has to be determined, evaluating how well they fit the original data. Finally the extracted factors are interpreted.

In our analysis the correlation matrix (Table 2) shows that all variables have at least one correlation coefficient with an absolute value larger than 0.6. We notice that percentage of households connected to the Internet is highly correlated with the number of computers, the percentage of active population using computers for professional purposes and the percentage of businesses buying on line, all with values over 0.8. Besides we must highlight the quite strong correlation between the percentage of businesses buying on line, the number of secure servers and the percentage of businesses with a website. It is also important to notice that, as expected, Internet access costs are negative correlated with all the variables, but one (broadband connections).

Table 2. Correlation matrix

	PC	TEL	BRO	SER	WEB	B2B	COST	HOH	PS	APOP
PC	**1.000**	**0.665**	0.572	**0.814**	**0.647**	**0.668**	-0.285	**0.863**	0.059	**0.812**
TEL	**0.665**	**1.000**	0.505	0.443	0.449	0.334	-0.023	0.481	-0.280	**0.686**
BRO	0.572	0.505	**1.000**	0.283	0.442	0.450	0.070	**0.650**	-0.301	**0.616**
SER	**0.814**	0.443	0.283	**1.000**	**0.662**	**0.705**	-0.330	**0.763**	0.191	**0.672**
WEB	**0.647**	0.449	0.442	**0.662**	**1.000**	**0.823**	**-0.606**	**0.796**	0.220	**0.782**
B2B	**0.668**	0.334	0.450	**0.705**	**0.823**	**1.000**	-0.495	**0.801**	0.338	**0.730**
COST	-0.285	-0.023	0.070	-0.330	**-0.606**	-0.495	**1.000**	-0.414	**-0.616**	-0.422
HOH	**0.863**	0.481	**0.650**	**0.763**	**0.796**	**0.801**	-0.414	**1.000**	0.255	**0.842**
PS	0.059	-0.280	-0.301	0.191	0.220	0.338	**-0.616**	0.255	**1.000**	-0.012
APOP	**0.812**	**0.686**	**0.616**	**0.672**	**0.782**	**0.730**	-0.422	**0.842**	-0.012	**1.000**

Note: Marked coefficients have an absolute value larger than 0.6

As the correlation matrix reveals the existence of strong relationships between some variables, this leads us to think that principal components analysis must be suitable.

Table 3 shows the results of principal components analysis in which the new components have been obtained through linear combinations of our set of ten variables. The components are ranked according to the share of explained variance. Thus Factor 1 explains almost 58% of the variation in the original variables; Factor 2 explains 20%; Factor 3, 6% and so on. In the end the ten factors explain 100% of variance.

As we want to reduce the complexity of our problem we have to decide the number of factors to extract. There are several criteria to do it: the eigenvalue criterion and the percentage of variance criterion, among others. The eigenvalue criterion states that all factors having eigenvalues greater than 1 should be retained. This guarantees that any factor accounts for at least the variance of a single variable. Meanwhile the percentage of variance criterion considers all factors accounting for at least 60% (typically)[4] of the variance of the original variables. In our case both criteria suggest retaining two factors, which explain almost 78% of total variance of the original indicators.

[4] Nevertheless, some authors suggest at least 80%.

Table 3. Results of principal components analysis

Factor	Eingenvalue	Percent of variance	Cumulative percent of variance
1	5.779	57.788	57.788
2	2.001	20.007	77.795
3	0.633	6.335	84.130
4	0.579	5.792	89.922
5	0.391	3.908	93.830
6	0.255	2.553	96.383
7	0.143	1.427	97.811
8	0.112	1.125	98.936
9	0.080	0.802	99.737
10	0.026	0.263	100.000

Note: Extracted factors are marked

Moreover the two underlying factors explain between 67% and 90% of the variance of each individual variable. This highlights the good quality of our results.

After the number of factors has been determined, the next step is to interpret them based on the matrix with the correlation coefficients between each factor and each variable (factor loadings matrix). To provide a more interpretable factor structure literature suggests the orthogonal rotation of this matrix so that each factor has high loadings on few variables. In particular we have applied the often-used varimax rotation, introduced by Kaiser in 1958. Table 4 shows the results.

Factor 1 is related to **ICT infrastructure** and **diffusion**, having positive loadings on all variables. We notice the significant positive loadings on the percentage of active population using computers for professional purposes, the number of computer and the percentage of households connected to the Internet, all over 0.9. It also represents the percentage of businesses with a website, the percentage of businesses buying on line as well as the number of secure servers, telephone lines and broadband connections.

Factor 2 is related to **e-government** and **Internet access costs**. It has positive loadings on the percentage of public services on line and negative loadings on Internet costs.

Table 4. Varimax rotated factor matrix

Variables	Factor 1	Factor 2
APOP	0.929	0.107
PC	0.911	0.085
HOH	0.901	0.287
WEB	0.793	0.419
B2B	0.765	0.460
SER	0.764	0.302
TEL	0.747	-0.325
BRO	0.741	-0.355
PS	-0.070	0.896
COST	-0.274	-0.823

Note: Variables are marked according to factor loadings

Denmark and Sweden are the countries with highest scores on Factor 1; meanwhile Portugal and Greece get the lowest. Concerning Factor 2, Finland and Luxemburg get the highest and the lowest scores respectively.

4.2. Cluster analysis

Using the two identified factors we have run a cluster analysis to look for groups of countries with similar levels of digital development.

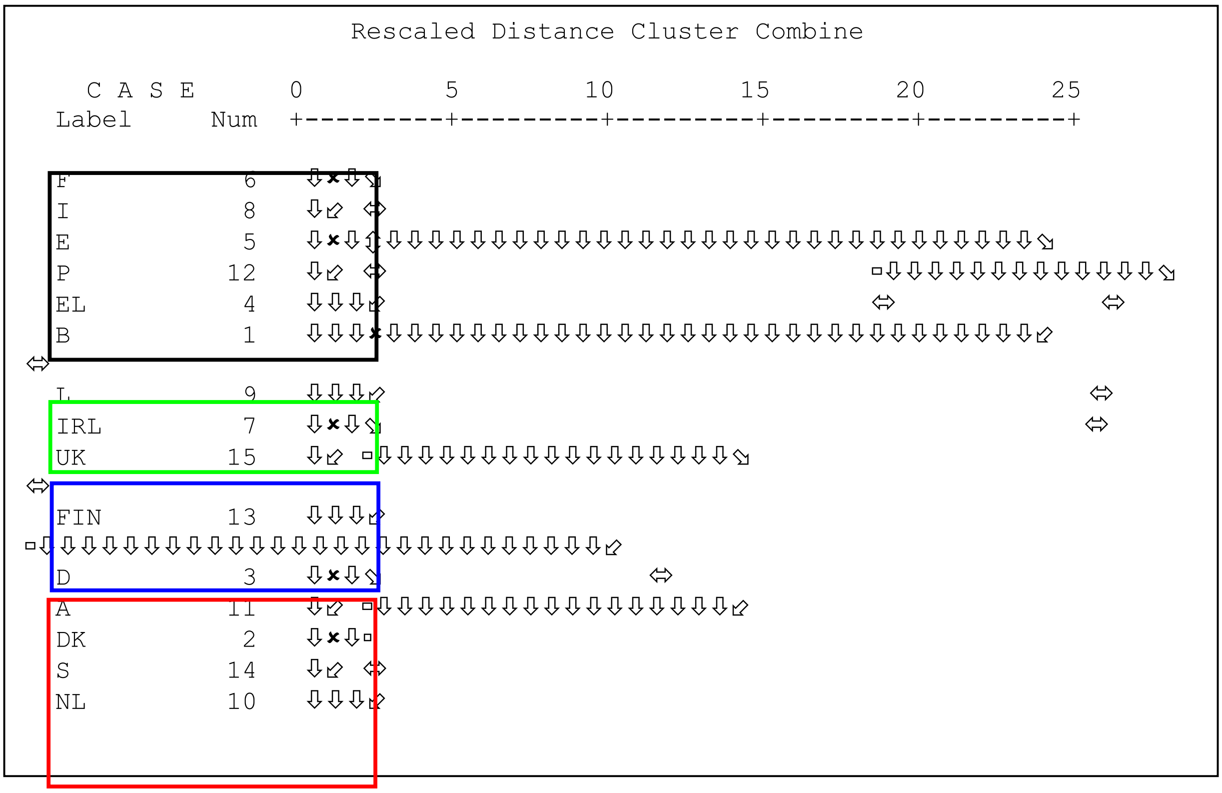
Cluster analysis techniques fall into two main categories: hierarchical and non-hierarchical methods. In hierarchical procedures, different numbers of clusters may be formed depending both on how similarity between observations is measured and on how groups are linked. On the contrary, non-hierarchical procedures start with a pre-determined number of clusters and then resorts to various algorithms to assign individual observations to different clusters.

To group European countries we have used the hierarchical approach as *a priori* we do not know the number of clusters that may be formed. Similarity between observations has been measured by means of the Squared Euclidean distance, as it is the often-used measure with quantitative data. Moreover we have tried different linkage rules.

Results from the majority of our analyses led us to consider the same four clusters. Besides, as we have run the cluster analysis with just two variables, results are quite easy to interpret.

A graphical representation of the solution is shown in Figure 1. The horizontal axis shows the linkage distance and the vertical axis arrays the countries identified by their abbreviations. Table 5 shows the principal differentiating characteristics of the four clusters.

Figure 1. Dendogram for the Digital Divide across European countries
(Squared Euclidean distance, Ward's linkage method)



Note: Ward's method.

Table 5. Descriptive statistics for the identified clusters

	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
Mean	-1.16	-0.19	0.58	-1.73	0.09	1.40	0.88	0.05
S.D.	0.41	0.47	0.82	0.23	0.48	0.48	0.47	0.37
N	5	5	2	2	3	3	5	5

Thus cluster 1 includes France and countries from Southern Europe. These are the less developed countries in terms of the ICT revolution as the average value for Factor 1 is -1.16, the lowest in the four clusters.

Cluster 2 consists of just Belgium and Luxemburg, which are the countries with the lowest provision of public services on line and the highest access costs. Notice that they get an average score on Factor 2 of -1.73.

Then cluster 3 includes Ireland, United Kingdom and Finland. Their main common feature is that they all have a high provision of public services on line as well as low Internet costs, contrary to cluster 2.

Finally cluster 4 consists of countries from Central and Northern Europe, in particular Germany, Austria, Denmark, Sweden and the Netherlands. These are the countries with the highest ICT penetration rates across the EU as they get the highest score average on Factor 1 with 0.88.

5. Concluding remarks

In this paper we have presented a new approach to the analysis of the digital divide by means of multivariate methods.

Principal components analysis has shown to be a useful technique in dealing with the multidimensionality of the digital gap. Thus it has allowed us to reduce the complexity of working with a set of indicators which are correlated with each other, as well as to identify the underlying dimensions of digital development across the EU (ICT infrastructure and diffusion; e-government and Internet access costs).

Then cluster analysis has allowed us to find four groups of countries with similar levels of ICT development. We have noticed that the well-known North-South divide reappears, but with the peculiarity that France joins the less developed group. We should also highlight the situation of Belgium and Luxemburg, where the rate of digital development is not accompanied by the provision of on line public services and the fall in Internet access costs.

All these results reflect the asymmetry in the development of the information society within the EU, and thus the need to reinforce policy actions in this area. Further research into this topic might explore digital disparities in the Acceding and Candidate countries, expanding the range of indicators.

6. References

- Annan, K. (2003): *Speech by UN Secretary General at the World Summit of the Information Society*, Geneva, (http://www.wsis-online.net/).
- Corrocher, N.; Ordanini, A. (2002): "Measuring the digital divide: a framework for the analysis of cross-country differences", *Journal of Information Technology*, 17, pp.9-19.
- Deutsche Bank (2001): "The Digital Divide: a focus on Asia", *Economics. Internet Revolution and new economy*, 17, Deutsche Bank Research, (http://www.dbresearch.de/).
- Eurostat (2002): *Statistics on the Information Society in Europe*, Eurostat, Brussels.
- Eurostat (2003): *Information society statistics pocketbook*, Eurostat, Brussels.
- Hargittai, E. (2003): "The Digital Divide and what to do about it", *New Economy Handbook*, Academic Press, New York, pp.821-839.
- Hilbert, M. R. (2001a): *From industrial economics to digital economics: an introduction to the transition,* Serie Desarrollo Productivo, 100, ECLAC, Santiago de Chile.
- Hilbert, M. R. (2001b): *Latin America on its path into the digital age: where are we?*, Serie Desarrollo Productivo, 104, ECLAC, Santiago de Chile.
- IDC/World Times: *Information Society Index*, (http://www.idc.com/).
- OECD (2001): *Understanding the Digital Divide*, OECD, Paris.

- OECD (2002): *The Digital Divide: Diffusion and use ICTs,* OECD, Paris.
- Oliveira, J.; Lourenço, M.M.; Ferreira. C.M. (2003): "A multivariate methodology to uncover regional disparities: A contribution to improve European Union and governmental decisions", *European Journal of Operational Research*, 145, pp.121-135.
- ORBICOM (2002): *Monitoring the Digital Divide*, (http://www.orbicom.uqam.ca/).
- ORBICOM (2003): *Monitoring the Digital Divide…and beyond*, (http://www.orbicom.uqam.ca/).
- PPI (1998): *The New Economy Index. Understanding America´s Economic Transformation*, Progressive Policy Institute, Washington.
- PPI (1999): *The State New Economy Index. Benchmarking Economic Transformation in the States*, Progressive Policy Institute, Washington.
- PPI (2001): *The Metropolitan New Economy Index. Benchmarking Economic Transformation in the Nation's Metropolitan Areas*, Progressive Policy Institute, Washington.
- PPI (2002): *The 2002 State New Economy Index. Benchmarking Economic Transformation in the States*, Progressive Policy Institute, Washington.
- Sciadas, G. (2002): "Unveiling the Digital Divide", *Connectedness Series,* 7, Statistics Canada.
- UN ICT Task Force (2003): *Tools for Development: Using Information and Communications Technology to Achieve the Millennium Development Goals*, (http://www.unicttaskforce.org/).
- UNDP (2001): *Making new technologies work for human development, Human Development Report 2001*, (http://hdr.undp.org/reports/).
- UNESCO (2003): *Measuring and monitoring the information and knowledge societies: a statistical challenge*, UNESCO, Paris.
- US Department of Commerce (1995): *Falling through the Net: A Survey of the 'Have-Nots' in Urban and Rural America*, (http://www.commerce.gov/).
- US Department of Commerce (1998): *Falling through the Net II: New Data on the Digital Divide*, (http://www.commerce.gov/).
- US Department of Commerce (1999): *Falling through the Net III: Defining the Digital Divide*, (http://www.commerce.gov/).
- US Department of Commerce (2000): *Falling through the Net IV: Toward Digital Inclusion,* (http://www.commerce.gov/).
- US Department of Commerce (2002): *A Nation Online: How Americans Are Expanding Their Use of the Internet*, (http://www.commerce.gov/).
- Vicente, M.R.; López, A.J. (2004a): "Disability in the Information Society: Facts and Challenges", *Anals of Applied Economics*, XVIII Asepelt-Spain Meeting, León, (in Spanish).
- Vicente, M.R.; López, A.J. (2004b): "Information and Communications Technologies in Latin America. Approaching the Digital Divide", subject to review.

- Warschauer (2002): “Reconceptualizing the Digital Divide”, *First Monday*; 7 (7), (http://www.firstmonday.dk/).
- World Economic Forum (2002): *Global Information Technology Report 01-02*, Oxford University Press, USA.
- World Economic Forum (2003): *Global Information Technology Report 02-03*, Oxford University Press, USA.
- World Economic Forum (2004): *Global Information Technology Report 03-04*, Oxford University Press, USA.
- World Summit on the Information Society (2003): *Building the Information Society: a global challenge in the new Millennium. Declaration of Principles*, Document WSIS-03/ GENEVA/DOC/0004, (http://www.itu.int/wsis/documents/doc_multi-en-1161|1160.asp).