

A multiple indicator model for panel data: an application to ICT area-level variation

Eva Ventura
Department of Economics and Business
Universitat Pompeu Fabra
and Barcelona GSE

Albert Satorra
Department of Economics and Business
Universitat Pompeu Fabra
and Barcelona GSE

May 2014

Abstract

Consider the case in which we have data from repeated surveys covering several geographic areas, and our goal is to characterize these areas on a latent trait that underlies multiple indicators. This characterization occurs, for example, in surveys of information and communication technologies (ICT) conducted by statistical agencies, the objective of which is to assess the level of ICT in each area and its variation over time. It is often of interest to evaluate the impact of area-specific covariates on the ICT level of the area. This paper develops a methodology based on structural equations models (SEMs) that allows not only the ability to estimate the level of the latent trait in each of the areas (building an ICT index) but also to assess the variation of this index in time, as well as its association with the area-specific covariates. The methodology is illustrated using the ICT annual survey data collected in the Spanish region of Catalonia for the years 2008 to 2011.

Keywords: structural equations model; confirmatory factor analysis; longitudinal analysis; index; digital divide; Information and Communication Technologies (ICT)

MSC classification: 62H25, 62-07, 91B82

1 Introduction

Statistical offices worldwide invest resources in surveys to ascertain the levels of information and communication technology (ICT) activity in the population. Often, the aim is to inspect the levels of ICT in different geographical areas rather than the levels of ICT of individuals. The level of ICT of an area may be important to determine its economic performance and competitiveness. That is, better ICT infrastructures and more skillful users of the digital technologies will result in improved business opportunities for the area. It is important therefore to measure the level of ICT activity for each area and to monitor the area and time variation of this index.

Many authors have studied the so-called digital divide across countries or regions. A commonly accepted definition of the term "digital divide" is provided by the OECD (2001). According to this organization, "the term digital divide refers to the gap between individuals, households, businesses and geographical areas at different socioeconomic levels with regard to their opportunities to access information and communication technologies and to their use

of the Internet for a variety of activities". The digital divide is thus, in nature, a multidimensional concept. In a recent paper, Vicente & López (2011) identify two main types of contribution to the literature of the digital divide. The first type focuses on "measuring and quantifying the extent, evolution, and pace of the digital divide", while the second tries to explain "the determinants of ICT disparities". There are a considerable number of studies that examine the relationship between the ICT disparities and several socioeconomic characteristics at the cross-country, regional and even the individual level. Interested readers may consult the work of Vicente & López (2011) for a complete overview. Echoing their work and regarding the measurement issue, there have been many attempts to build composite indices capable of capturing the "multidimensionality of the digital divide" (Al-Mutawkkil, Bagchi, 2005; Corrocher & Ordanini, 2002; Cruz-Jesús, Oliveira & Bacao, 2012; Dutta & Jain, 2004; Heshmati & Hwang, 2009; ITU, 2009; Orbicom, 2003; Selhofer & Hüsing, 2002; Wolcott et al., 2001). Simple compound indices based on the arbitrary weighting of several indicators present certain disadvantages. New technologies are continuously emerging and have to be incorporated into the calculation of the index. Therefore, it is necessary to constantly update both the definition of the index as well as the weights used to calculate it. On the other hand, the weights are arbitrary and "could be the subject of political dispute" (Cruz-Jesús, Oliveira & Bacao, 2012). Thus, a number of these studies rely on exploratory factor analysis to select the weights applied to a number of indicators to produce a single composite index (Al-Mutawkkil et al., 2009; Corrocher & Ordanini, 2002; Cruz-Jesús, Oliveira & Bacao, 2012; ITU, 2009; Hanafizadeh, Saghaei & Hanafizadeh, 2009; Soupizet, 2004; Vicente & López, 2011). Other authors use multivariate techniques such as multivariate analysis, cluster analysis and discriminant analysis to "explore the digital divide among old, new and candidate member states of the European Union" (Çilan et al., 2009; Cruz-Jesús, Oliveira & Bacao, 2012). In this line, Vehovar et al. (2006) argue that the multivariate approaches such as loglinear modeling, compound measures and time-distance methodology used to analyze the changes in the digital divide are a better alternative to "an oversimplified methodological approach to digital divide studies".

Much of the work performed on measuring and explaining the digital divide has been performed at the cross-country level. Research at the regional level has been conducted in large part for the United States (Atkinson & Andes, 2008; Chaudhuri, Flamm & Horrigan, 2005; Grubestic, 2006; Horrigan, Stolp & Wilson, 2006; Kolko, 2000; Mills & Whitacre, 2003, Progressive Policy Institute, 2002; US Department of Commerce, 2000); the European references are fewer (Billón, Ezcurra & Lera-López, 2008, 2009; Vicente & López, 2007, 2011). There are a few studies that use micro data to determine the urban/rural digital gap (Chaudhuri et al., 2005; Mills & Whitacre, 2003; Noce & McKeown, 2008; US Department of Commerce, 2000), and others that use the individual and regional characteristics to explain the digital divide (Demoussis & Giannakopoulos, 2006; Horrigan et al., 2006; Schleife, 2010).

This paper presents a methodology that integrates in a single model both the measurement of the ICT activity for each area in each period, as well as the association of the ICT activity with the area-specific covariates. As we have already mentioned, the idea of an index based on factor analysis is not new. The approach in this paper is unique because the confirmatory factor analysis replaces the principal components approach. This permits us to assess the dimensionality of the index and to study its evolution in time for each of the areas. The geographical areas are meant to be the territories that differ in the level of contingency of several factors that we feel explain the digital gap among the areas. In previous works based on the principal components and exploratory factor models, the estimated factors are regressed into a series of explanatory variables to determine what may cause the differences among the regions or countries (see, for instance, Vicente & López, 2011). In contrast to this two-stage approach (fitting first an exploratory factor model for each year and then applying a regression to the extracted factors), our methodology considers the single-stage analysis for

which the level of ICT activity is estimated and simultaneously regressed on the covariates. This single-stage approach leads to a more consistent and efficient estimation.

To illustrate our methodology, we use the ICT data collected in the administrative areas of the Spanish region of Catalonia. The Statistics Institute of Catalonia (Idescat) conducts repeated household surveys with the aim of ascertaining the time variation in the ICT at the area level. We combine several survey questions to produce a set of indicators representing the ICT activity at the individual level. These variables are then aggregated within each area to produce the indicators of the ICT activity at the area level.

The rest of the paper is organized as follows. Section 2 describes the data to be used. A confirmatory factor analysis with one factor for each year is developed in section 3. Section 4 presents a panel data model whereby the factors are decomposed into a permanent and a time-varying component and related to the covariates. Section 5 presents the conclusions.

2 Data and Construction of Indicators

In this application, we use a household survey on the ICT conducted by the Idescat. The survey is a rotating panel, with five shifts of rotation on an annual basis. Here, we use four consecutive surveys, from 2008 to 2011. The sample design is by means of a stratified two-stage sample. The strata are 41 Catalan administrative divisions, and the sample is uniformly distributed among them with 75 randomly-selected first-stage units (dwellings) in each administrative division. Within each dwelling, the second-stage unit is one person aged 16 or over (also randomly selected).

Each one of the surveys asks more than fifty questions, distributed among several thematic blocks such as the main home equipment in terms of the ICT products, home access to the Internet, satisfaction of the respondent¹, use of computers by the respondent, use of Internet by the respondent, e-commerce, language and ICT, the knowledge and use of public Internet access points, trust in digital services, and courses and computer skills. Many of the answers to those questions provide us with dichotomous indicators that are highly correlated to each other. Other indicators attain values on a discrete scale. Some of the questions are redundant, and the surveys also collect information on certain socioeconomic issues of the households.

The first step consists of identifying the underlying factors that determine the level of ICT equipment and/or ICT-related skills of the household. This issue has been explored by many academic articles that address the measuring of the digital divide, and we will rely on their findings. Corrocher & Ordanini (2002) refer to the model elaborated by the OECD Task Force on the digital economy (Colecchia, 2000) and consider three stages of digital development. At the beginning, the factors that determine the differences between countries or regions are related to the speed of adoption of the new technologies. There are three main factors: the communication infrastructures, human resources and competitiveness of the companies providing the information and communication technologies. Subsequently, the technology spreads out though the territory, and the aspects related to the intensity of the adoption of the technology become more important. Here, Corrocher & Ordanini (2002) introduce the market diffusion factors, meaning "the diffusion of different devices for the use of digital services and applications that can determine different patterns of digitalization in different systems" and the size of the digital market. In the third stage, the qualitative aspects of the ICT become more relevant. The factors to consider are related to the impact of digitalization on the social and economic activities, on the structure of production, or on employment.

Our data set contains some indicators that can be used to measure several of those factors. We will aggregate and summarize the information they provide into several new variables, which in the tradition of the factor analysis approach will be used as indicators of a

¹ These are a group of questions dealing mainly with the number and type of incidences related to the Internet services provided by the companies in the digital sector.

common factor of the ICT activity for individuals. The set of indicators is essentially stable along the different years, though there are some minor variations that can be addressed by our methodology.

The second and third blocks of the survey questions contain information on the equipment and access to the Internet in the household. The information is summarized in a new variable (*equip*) that takes values ranging from 0 to 3. Households without a personal computer and without access to the Internet score 0 in this variable, while those with a personal computer and broadband access score 3. Other intermediate situations score 1 or 2 depending on the quality of the equipment and access to the Internet. The two blocks also contain information on the number of operating mobile phones in the household (*mobile*), and in 2008 we also know the length of time that there has been Internet access at the home² (*sinceInt*). This last variable is important to determine the degree of the ICT preparedness of the household and will be correlated with the respondents' ICT skills and with the measures of the social and economic impact of the ICT. It ranges from 0 if the household does not have access to the Internet to 4 if there has been access for over five years.

As for the measurement of human resources, we consider several indicators. The survey systematically records all the abilities of the respondents using a computer and/or the Internet. There are seven questions on the tasks performed with a computer (such as the management of files and documents) and also nine questions on the uses of the Internet. We then create new variables (*knowPC* and *knowInt*), which add together the number of distinct tasks that the respondent is capable of doing without the help of others, either with a personal computer or through the Internet. The variables range from 0 for the respondents that do not own a computer or do not have access to the Internet to 7 for the respondents with high computing skills and 9 for the respondents with proven abilities using the Internet.

We also take into account the intensity of the use of a personal computer or the Internet. The indicators measuring the intensity of use will most likely be correlated with the knowledge variables. The indicator that measures the frequency of use of the home computer (*freqPC*) takes values ranging from 0 to 5. The value is 0 if there is no computer in the household and 5 if the computer is used daily. As for the frequency of use of the Internet, the newly created variable (*freqInt*) ranges from 0 to 7. Again, the value is 0 if the household does not have access to the Internet, and it is 7 if the respondent is connected to the Internet for more than 20 hours a week.

On the chapter regarding the market diffusion factors, the household survey has little to offer. We have chosen to include the indicators related to the perceived level of security in three types of electronic transactions. The survey asks the respondent to what extent she or he thinks it is safe to use a credit card in a restaurant, to buy through the Internet, or to perform bank transactions through the Internet. Further analysis revealed that the security perception when paying in a restaurant with a credit card is only weakly correlated with the rest of the indicators. Therefore, we did not include that variable in the set of newly created variables. The answers to the last two security perception questions are translated into two new variables (*secbuy* and *secbank*) with values that range from 0 for a perception of total insecurity to the value of 3 if the respondent declares feeling very secure.

Finally, we create a number of synthetic indicators aimed at capturing part of the social and economic impact of the ICT. The social and economic benefits of these technologies should be more evident for the frequent users of the Internet. The variable created to measure the economic impact of the ICT (*econ*) ranges from 0 to 4. The value is higher for the frequent users performing several tasks of an economic nature, such as telework or buying through the Internet. The survey also records several types of activities -most of them related to leisure-

² This question was not asked in subsequent surveys.

that can be viewed as increasing the social exposure of the respondents. For instance, there are questions on the participation in chats, social networks and the use of virtual communities. We sum the answers to these questions to create a new variable (*social*) that ranges from 0 to 4. The level of interaction with several public health administrations is also measured by combining the response to several survey questions. The new variable (*admin*) ranges from 0 to 6.

Table 1 summarizes these synthetic indicators.

Table 1: Set of synthetic indicators

Variable	Description	Values			
		2008	2009	2010	2011
<i>econ</i>	Level of equipment in the household	0 to 3	0 to 3	0 to 3	0 to 3
<i>mobile</i>	Number of mobile phones in the household	0 to 5	0 to 5	0 to 5	0 to 5
<i>sinceInt</i>	Since when has access to Internet	0 to 4	--	--	--
<i>knowPC</i>	Number of tasks performed with a computer	0 to 7	0 to 7	0 to 7	0 to 7
<i>knowInt</i>	Number of tasks performed through Internet	0 to 9	0 to 9	0 to 9	0 to 9
<i>frecPC</i>	Intensity of use of personal computer	0 to 5	0 to 5	0 to 5	0 to 5
<i>frecInt</i>	Intensity of use of Internet	0 to 7	0 to 7	0 to 7	0 to 7
<i>secbuy</i>	Security perception: buying through Internet	0 to 3	0 to 3	0 to 3	0 to 3
<i>secbank</i>	Security perception: bank transactions through Internet	0 to 3	0 to 3	0 to 3	0 to 3
<i>econ</i>	Level of economic impact of ICT actions from home	0 to 4	0 to 4	0 to 4	0 to 2
<i>social</i>	Level of social impact of ICT actions from home	0 to 4	0 to 4	0 to 6	0 to 3
<i>admin</i>	Level of interaction with public administrations through Internet	0 to 6	0 to 6	0 to 7	0 to 7

A final linear transformation was applied to all the indicators, so they range from 0 to 10. Because we were interested in estimating the level of ICT activity at the area level, we used the sampling weights included in the database to compute the areas' sample averages involved in the posterior analysis.

3 A Confirmatory Factor Model for Longitudinal Data

We apply a confirmatory factor analysis (CFA) in which the level of the ICT activity of a particular year is a factor (latent variable) underlying the set of indicators of the ICT activity at the area level listed in Table 1.

Formally, we consider one factor³ in each of T consecutive periods. Areas are indicated by subscript $i = 1, 2, \dots, n$, and there are K synthetic indicators measuring each factor⁴. The measurement equations of the CFA model are:

$$y_{itk} = \lambda_{ik} f_{it} + e_{itk} \quad t = 1, 2, \dots, T; \quad k = 1, 2, \dots, K \quad (1)$$

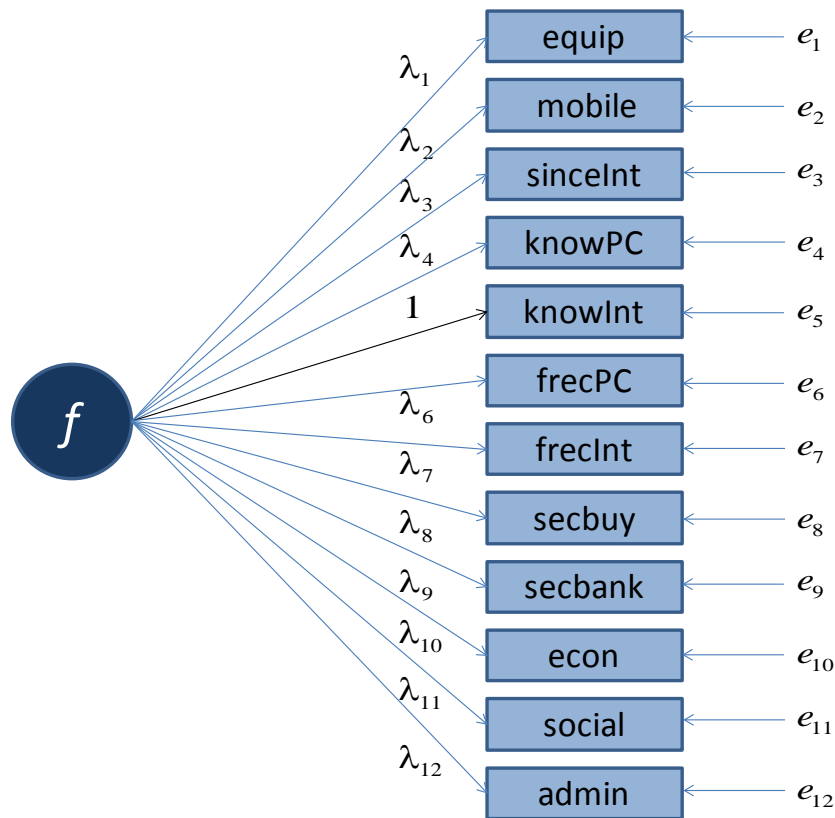
³ If necessary we could have introduced more than one factor for each period. In our application, a single-factor model led to an acceptable fit.

⁴ In our model, the 2008 factor has twelve indicators, while the rest of the factors have only eleven. For the sake of simplicity, we do not take this detail into account.

Here, y_{itk} is the k th indicator of the factor at time t for the i th area, f_{it} is the factor (level of ICT activity) of area i at time t , and λ_{ik} is the factor loading of the k th indicator at time t . The errors e_{itk} are uncorrelated with each other and with the factors.

The measurement part of the model for year 2008⁵ is depicted as a path diagram in Figure 1. Rectangles represent the observed indicators, and the factor f is in a circle. There are arrows going from the factor to the indicators, representing the loadings (λ_k). To fix the scale of the factors, the loading of variable *knowInt* (Internet skills) is set to one. There are also arrows going from each disturbance term e_k to the corresponding indicator. A similar path diagram applies for the other years.

Figure 1: Path diagram of the 2008 section of the CFA model

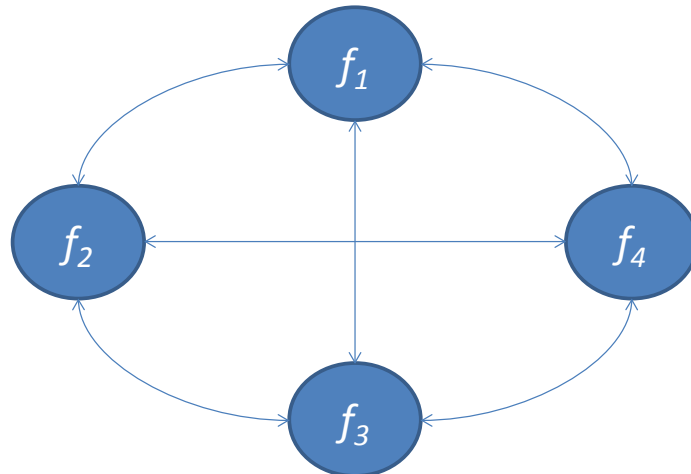


Allowing correlation parameters among the factors ties the four path diagrams into a single CFA model. Figure 2 displays the path diagram for the four correlated factors whereby a double-headed arrow represents a correlation. This is a confirmatory factor model because the indicators of one year have zero loadings on the factors of other years. The estimation of this CFA model is undertaken using EQS (Bentler, 2002)⁶. Other SEM software, e.g., LISREL, Mplus, CALIS, *sem* of Stata, AMOS, etc. could have been used.

⁵ This measurement part is almost identical every year; the only difference is that one of the indicators (*sinceInt*) is only present in year 2008.

⁶ In EQS, we use the least squares (LS) estimation option with robust standard errors. For the technical details of this method, see Satorra & Bentler (1990, 1994).

Figure 2: Path diagram of the correlated factors



The estimation results for this CFA model⁷ are displayed in Table 2 and Table 3. The first table reports the estimates of the factor loadings for the full model. Each column corresponds to the factor loadings of one year, with the factors designated f_1 in 2008 to f_4 in 2011. The second table displays the estimated correlation matrix of the factors.

All the factor loadings estimates are statistically significant and positive. The indicator with the highest loading is *frecInt* (frequency of use of the Internet) followed by *frecPC* (frequency of use of personal computer) and *equip* (equipment and access to Internet in the home). The relevance of indicator *knowPC* (personal computer skills) grows with time, with factor scores of 0.795 in 2008 and 1.179 in 2011. In addition, the indicator *social* (the use of ICT to engage in social activities related to leisure) is associated to a high (and increasing with time) factor loading. The two indicators related to security perceptions (*secbank* for the bank transactions by Internet and *secbuy* for electronic commerce) exhibit lower values for their respective factor loadings. The value for the factor loading is also low for the variables measuring the economic impact (*econ*) and the interaction with public administrations (*admin*) when using the ICT. The number of mobile phones in the home (*mobile*) seems to be moderately important in the year 2008 (a factor loading of 0.564) but becomes less important in consecutive waves. This is consistent with the widespread implementation of such devices and therefore the indicator *mobile* becomes less important to explain the variation in the level of the ICT activity across the areas.

⁷ We display the estimated results without imposing an equality constraint on the loadings across time. The same model could be estimated with the loadings set as equal across time. Such a model would be an alternative specification because the robust (mean- and variance-adjusted) chi-square goodness of fit test yielded a nearly acceptable fit (Chi-square = 25.627 with 15 degrees of freedom, p -value = 0.042).

Table 2: Factor loadings estimates in the CFA model

Indicator	f_1	f_2	f_3	f_4
	2008	2009	2010	2011
<i>equip</i>	1.084 (0.157)	1.187 (0.197)	1.136 (0.156)	0.793 (0.123)
<i>mobile</i>	0.564 (0.099)	0.589 (0.096)	0.355 (0.152)	0.231 (0.092)
<i>sinceInt</i>	0.910 (0.123)	--	--	--
<i>knowPC</i>	0.795 (0.111)	0.978 (0.041)	1.129 (0.064)	1.179 (0.071)
<i>knowInt</i>	1	1	1	1
<i>freqPC</i>	1.035 (0.207)	1.174 (0.114)	1.109 (0.074)	1.039 (0.113)
<i>freqInt</i>	1.158 (0.059)	1.152 (0.067)	1.302 (0.050)	1.151 (0.080)
<i>secbuy</i>	0.527 (0.094)	0.405 (0.102)	0.344 (0.096)	0.511 (0.092)
<i>secbank</i>	0.215 (0.038)	0.135 (0.037)	0.149 (0.032)	0.160 (0.037)
<i>econ</i>	0.532 (0.088)	0.604 (0.071)	0.731 (0.072)	0.691 (0.126)
<i>social</i>	1.004 (0.066)	1.012 (0.097)	1.246 (0.063)	1.304 (0.143)
<i>admin</i>	0.325 (0.064)	0.378 (0.053)	0.399 (0.058)	0.377 (0.062)

Mean- and variance-adjusted goodness-of-fit model test: $\chi^2=30.746$, d.f. = 22 (p – value = 0.101).

Robust standard errors are displayed below the estimates, in parenthesis.

As observed in Table 3, the four factors are highly correlated, suggesting a very stable level of ICT activity across the years at the area level.

Table 3: Estimated correlation matrix of the factors (CFA model)

	f_1	f_2	f_3	f_4
f_1	1			
f_2	0.879	1		
f_3	0.807	0.802	1	
f_4	0.730	0.610	0.780	1

The empirical best linear unbiased estimators (EBLUP) of the factors were computed. These are the factor scores provided by EQS when the *regression* method is used (see Satorra & Bentler, 2014). These factor scores are displayed in Table 4 where the rows correspond to the areas and the columns correspond to years. The score for each area in each of the four columns corresponds to an estimate of the ICT level of activity for the corresponding year. These area scores are relevant for ranking the areas according to their ICT activity, as well as to assess the ICT variation across time. At this point, however, any ranking of areas is problematic because we have four scores for each area.

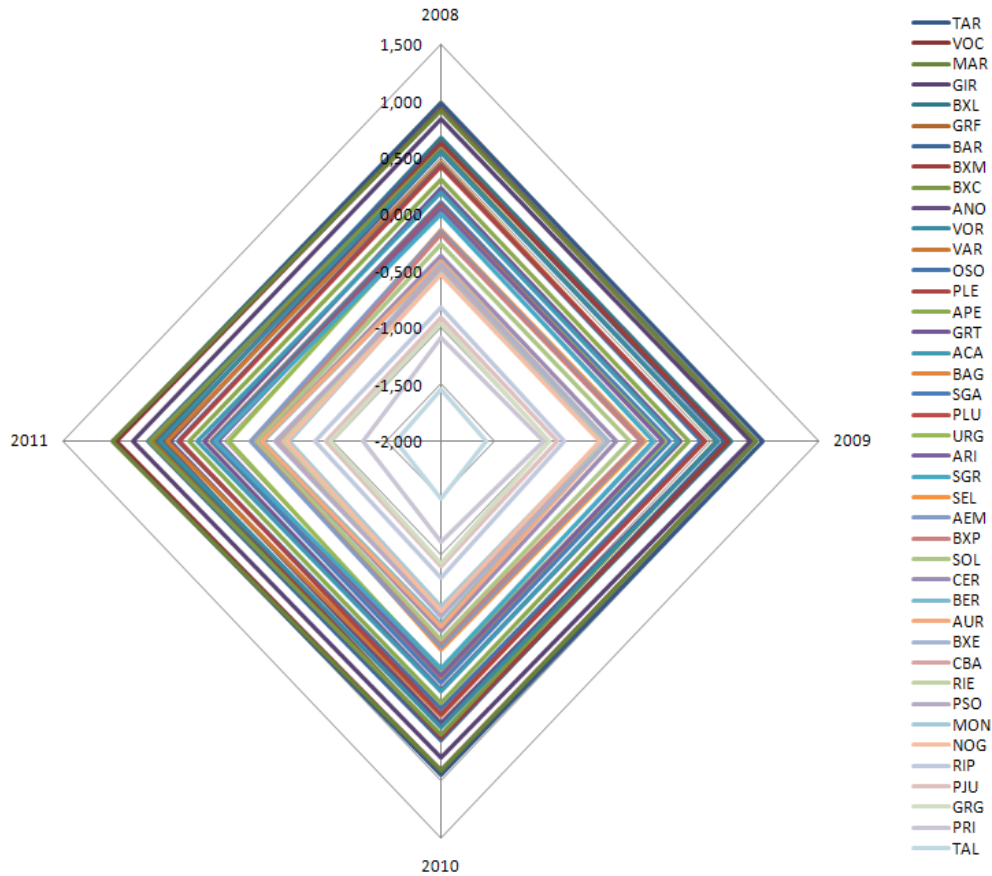
Table 4: Factor scores estimates (ICT index by area and time)

AREA	f_1	f_2	f_3	f_4
	2008	2009	2010	2011
TAR	0.982	0.985	0.947	1.031
VOC	0.922	0.917	0.909	1.000
MAR	0.913	0.912	0.908	1.038
GIR	0.840	0.866	0.799	0.858
BXL	0.673	0.696	0.614	0.632
GRF	0.639	0.647	0.618	0.672
BAR	0.630	0.629	0.646	0.719
BXM	0.625	0.665	0.611	0.577
BXC	0.572	0.553	0.587	0.703
ANO	0.551	0.560	0.495	0.514
VOR	0.549	0.558	0.533	0.597
VAR	0.450	0.429	0.422	0.534
OSO	0.427	0.413	0.369	0.436
PLE	0.424	0.445	0.415	0.425
APE	0.301	0.297	0.316	0.336
GRT	0.227	0.225	0.202	0.198
ACA	0.186	0.190	0.211	0.261
BAG	0.101	0.089	0.096	0.131
SGA	0.099	0.113	0.130	0.119
PLU	0.079	0.064	0.050	0.079
URG	0.050	0.078	0.033	0.034
ARI	0.047	0.057	0.075	0.066
SGR	-0.005	-0.032	0.007	0.085
SEL	-0.144	-0.108	-0.159	0.293
AEM	-0.152	-0.140	-0.186	0.239
BXP	-0.183	-0.152	-0.226	0.307
SOL	-0.273	-0.244	-0.243	0.305
CER	-0.372	-0.376	-0.320	0.341
BER	-0.416	-0.455	-0.388	0.348
AUR	-0.422	-0.450	-0.366	0.339
BXE	-0.450	-0.449	-0.445	0.515
CBA	-0.481	-0.506	-0.465	0.503
RIE	-0.502	-0.499	-0.488	0.538
PSO	-0.504	-0.521	-0.466	0.473
MON	-0.521	-0.521	-0.535	0.595
NOG	-0.530	-0.527	-0.506	0.565
RIP	-0.824	-0.852	-0.791	0.840
PJU	-0.909	-0.925	-0.889	0.939
GRG	-0.968	-0.997	-0.920	0.976
PRI	-1.092	-1.052	-1.113	1.284
TAL	-1.545	-1.581	-1.485	1.578

Note: areas have been ordered in decreasing values (top to bottom) of the row average.

To display the time evolution of the ICT activity for each area, the factor scores of Table 4 are represented as a radial plot in Figure 3. We have four radial axes, one for each year; the area axis coordinate is the factor score of the year. A polygon is obtained for each area when connecting the four coordinates. In our application, we observe that the polygons rarely cross each other, revealing a high stability across time for the ICT activity (in this application, this can be readily observed by inspecting the column variation in Table 4).

Figure 3: Radial plot of the factor scores



We observe that in the outskirts there are four areas (TAR, VOC, MAR and GIR) slightly ahead of the rest in terms of the ICT activity (the factor scores are over 0.8 for each year). These are dynamic and heavily urbanized areas, the economic activity mainly based on industry and services. At the other extreme in the center of the plot, there are five areas (RIP, PJU, GRG, PRI and TAL) that are at the bottom of ICT activity (the factor scores are below -0.79). These are rural, low density areas with an ageing population.

Two main results emerge from this CFA analysis. First, we observe a high variability in the ICT level across areas for each of the years. It is observed that in general, the urban, dynamic, and industrialized areas tend to be at the top of the classification, while the rural, agricultural and ageing areas are at the bottom. Second, there is little variation of ICT activity over time. The results suggest the need to expand the CFA model to the one developed in the next section.

4 A Multiple Indicator Panel Data Model with Covariates

We now propose a panel data model in which the ICT factors underlying the multiple indicators (f_1 to f_4) will be decomposed into two components: a permanent component and a transient one, with the permanent component being regressed on the observable socio-economic characteristics of the areas (covariates). An autoregressive structure is assumed for the time-varying component.

The model is expressed by the following equations. First, we have the measurement equation (1) of the previous section that relates the multiple indicators with the factors. Second, factors (f_{it}) are decomposed into permanent (p_i) and transitory components (q_{it})

$$f_{it} = p_i + q_{it} \quad (2)$$

with $i = 1, 2, \dots, n$ and $t = 1, 2, \dots, T$. In addition, we specify an autoregressive structure for the transitory component

$$q_{it} = \rho q_{it-1} + v_{it} \quad (3)$$

where v_{it} is a disturbance term uncorrelated with q_{it-1} . Finally, we have the regression equation

$$p_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi} + \delta_i \quad (4)$$

where the permanent component p_i depends on m area-level covariates x_j , $j = 1, 2, \dots, m$.

The full set of equations (1) to (4) is a panel data model in which the dependent variables are the common factors developed in the previous section. See Bou & Satorra (2007, 2014) for an application of this model to finance the data and technical details on this class of panel data models.

In our application, the permanent component p_i is likely to be substantial given the high correlation among the factor scores observed in the previous section (see Table 3). The practical relevance of this component will be determined by its variance. At one extreme, we could have a case in which the whole variance of f_{it} in equation (2) is captured by p_i (all the variation across the areas is permanent); and, at the other extreme, we could have a case in which the variance of p_i is negligible compared with the variance of the transitory component, so the relevant variation is fully captured by the transitory component.

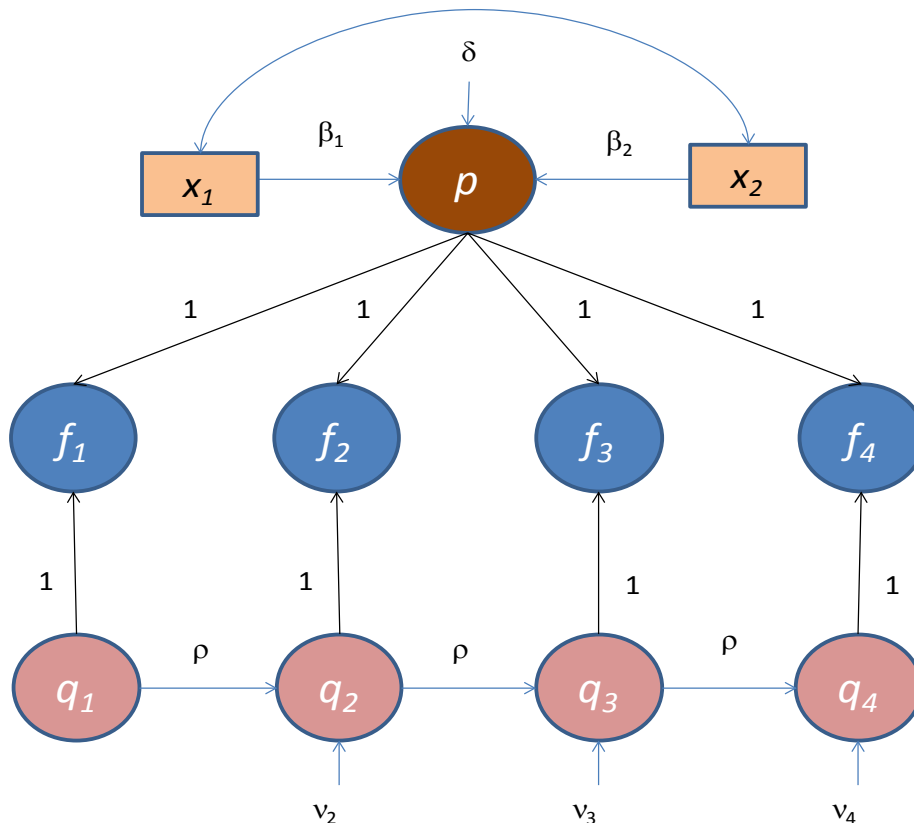
The autoregressive structure of the transitory component q_{it} in equation (3) simply assumes that the time-dependent ICT activity in an area has some persistence but vanishes with time. The speed of the time erosion of this component is given by the autoregressive parameter ρ . When ρ is zero the temporary component is independent across the years, while a lasting effect arises when ρ approaches one.

The choice of the covariates of equation (4) is determined by the availability of the socioeconomic data at a highly disaggregated level. According to the previous contributions on this issue, the likely causes of the observed disparities in the ICT level were the differences in the average regional economic development, productive structures (Billon et al. 2008, 2009), human capital (Demoussis & Giannakopoulos, 2006), demography (Holloway, 2005), investment in research and development (Vicente & López, 2006, 2011), cultural and institutional factors (Howard et. al, 2009), or the access costs to the ICT equipment and services (Kiiski & Pohjola, 2002). The Idescat publishes data at the administrative-unit level on the gross domestic product (GDP), the percentage of gross value added (GVA) by the type of economic activity, the percentage of occupied workforce by the type of economic activity, the unemployment rate, the population share by age, the educational level or the population density. However, information on the investment research and development and the cultural and institutional factors is not available.

Table B.1 in appendix B exhibits the correlations between the available set of variables (measured in 2008 or earlier) and the ICT activity at the area level. A goodness-of-fit

assessment of the alternative SEM specifications led us to the choice of the two covariates for our panel data model, namely, the percentage of younger population (variable x_1) and the percentage of gross value added in the agricultural sector (x_2). Figure 3 displays the path diagram corresponding to equations (2) to (4) of the panel data model. The measurement part of the model is identical to that of the CFA model except that we have equated the loadings across time. We also equated across time the variances of the disturbance terms of equation (3)⁸.

Figure 4: Path diagram representation of factors and covariates



This multiple-indicators panel data model is estimated with the EQS under LS and robust statistics. Table 5 displays the corresponding parameter estimates and standard errors (robust). The regression coefficients in equation (4) are both statistically significant, with variable x_1 (the share of the young population) positively affecting the permanent part of the ICT activity and variable x_2 (the share of agricultural activity) affecting it negatively. The determination coefficient R^2 for this equation is 0.82, revealing that a high percentage of the variation of the permanent component of the ICT activity is explained by the covariates.

⁸ Both sets of restrictions have been validated by goodness-of-fit testing.

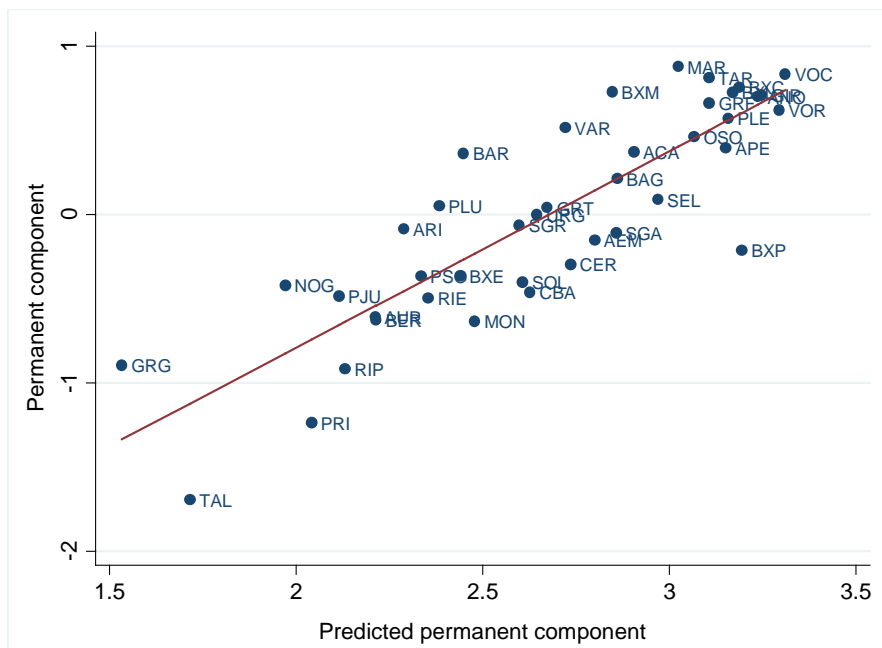
Table 5: Results for the panel data model.

Parameter	Estimate	Robust S.E.
<i>equip</i>	1.013	0.107
<i>mobile</i>	0.385	0.074
<i>sinceInt</i>	0.948	0.135
<i>knowPC</i>	1.007	0.040
<i>knowInt</i>	1	--
<i>frecPC</i>	1.119	0.087
<i>frecInt</i>	1.135	0.046
<i>secbuy</i>	0.386	0.060
<i>secbank</i>	0.160	0.017
<i>econ</i>	0.635	0.050
<i>social</i>	1.134	0.061
<i>admin</i>	0.368	0.044
β_1	0.183	0.032
β_2	-0.047	0.015
ρ	0.371	0.0146

Mean –and-variance goodness-of-fit model
test: $\chi^2=11.813$, d.f.= 11 (*p* – value = 0.378)

The factor scores (EQS, *regression* method) for both the permanent and temporal components of ICT activity are displayed in table C.1 of appendix C. A scatter plot of the estimates of the permanent component versus its predicted value by equation (4) is displayed in Figure 4. This graphic visualizes the ordering of the areas and their deviations from the predicted values of the (permanent) ICT activity given the covariates. For example, we observe that area BXP has a much lower ICT activity than was predicted. The inspection of the deviations of the areas from the group pattern could suggest issues for further research.

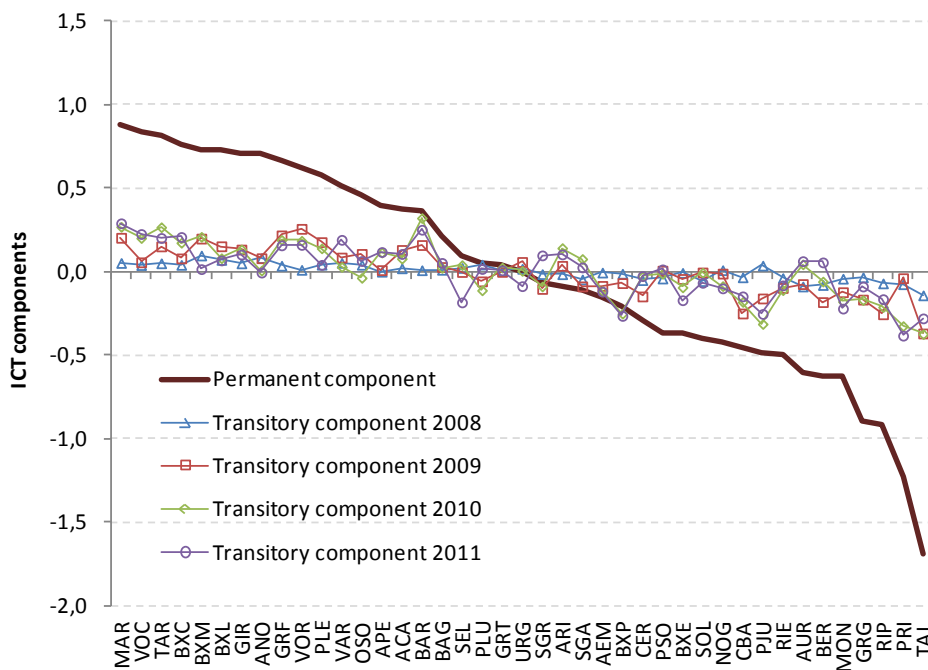
Figure 5: Predictors of ICT activity



The autoregressive coefficient ρ of equation (3) is significant but not very large, indicating that yearly variations in the ICT activity tend to vanish quickly in subsequent years. An important result of the analysis is that most of the variance (across areas) of the ICT level is explained by the permanent component. The standardized solution of equation (2) indicates that approximately 70% of the variation in the level of ICT activity is accounted for by the permanent component. See appendix D for the estimated equations (2) to (4) of the standardized model. Because the covariates explain a large percentage of the permanent component and the permanent component is a dominant part of the ICT variation, we conclude that the covariates highly determine the ICT variation. This issue could not have been detected by simply observing the correlations among the indicators and covariates.

Figure 5 provides a graphical representation of the relative size of the permanent and the temporary components of the ICT activity. The vertical axis of the graph corresponds to the values of the permanent or transitory components, and the horizontal axis contains the 41 areas ordered from the higher to the lower values of the permanent component. The solid line that crosses the graph from the value 0.88 for MAR to the value -1.694 for TAL corresponds to the permanent component. The four other lines correspond to the four temporary components, and we observe a much narrower range of variation.

Figure 6: Permanent and transitory components of ICT activity



5 Concluding Remarks

We presented a panel data methodology in which a common factor underlying a set of indicators of each time period is decomposed into the permanent and transitory components, and the permanent component is regressed on the covariates. The methodology uses standard SEM software and is applied to the data from repeated surveys on ICT activity in the districts of the region of Catalonia, Spain.

The analysis indicates a high degree of persistence in the level of ICT activity (with small variation over time) with the permanent component accounting for most of the ICT variation. A regression equation of the model demonstrates that approximately 82% of the variation of

this permanent component is accounted for by the variation in the observable socio-economic characteristics of the areas. The factor scores corresponding to the permanent component can be interpreted as an index of the ICT activity at the area level and thus allow the areas to be ranked. Deviation of this index from the value predicted by the covariates permits the assessment of the differential performance of the areas. In the context of our data, the marginal variance of the transient component (compared with the permanent component) raises the substantive questions of whether it is necessary for Idescat to repeat a costly ICT survey every year.

The modeling approach proposed in this paper could apply to other areas of official statistics for which surveys are periodically repeated, with multiple indicators tapping a latent construct that is the focus of interest (in our case, the level of ICT activity). The method integrates into a single model the multiple stages of analysis that are commonly performed using exploratory or dimension-reduction techniques, such as principal components, discriminant analysis, and other multivariate methods. Integrating into a single framework the different stages obviously leads to a more accurate and informative analysis. In our single-stage approach, we are able to neatly define the factor underlying the indicators, even to assess its dimensionality, and enable the regular test statistics to detect the structural changes in the model that are likely to arise in the applications. In addition, in our approach we obtain an overall goodness of fit test to validate our model. For applied researchers, it is important to note that our analysis can be routinely implemented using standard SEM software.

In our application, debate could arise at the stage of defining the multiple area-level indicators that have been used. However, for the methodological emphasis of the paper, we feel they are sufficient indicators of ICT activity. A requirement of our model is a sufficiently large sample size to ensure a negligible sample-survey variance of the indicators at the area level. In addition, it is necessary to incorporate a sufficient number of areas to implement the inferences of the SEM analysis. See Satorra & Bentler (2014) for corroboration that our data set-up fulfills these requirements and for the alternative methods to be used when the sample size in each area is small.

Acknowledgments

The authors would like to thank the Statistics Institute of Catalonia (Idescat) for facilitating the data used in this study. In addition, the authors wish to acknowledge the financial support from the Spanish Ministry of Science and Innovation through grant ECO2011-28875 .

References

- [1] Al-Mutawkkil, A., Hesmati, A., & Hwang, J. (2009). Development of telecommunication and broadcasting infrastructure indices at the global level. *Telecommunications Policy*, 33(3-4), 176-199.
- [2] Atkinson, R., & Andes, S. (2008). The 2008 state new economy index: Benchmarking economic transformation in the states. Retrieved from <<http://ssrn.com/abstract=1323828> or <http://dx.doi.org/10.2139/ssrn.1323828>>
- [3] Bagchi, K. (2005). Factors contributing to global digital divide: Some empirical results. *Journal of Global Information Technology Management*, 8(3), 47-65.
- [4] Bentler, P. M. (2002). EQS 6 Structural equations program manual. . Encino, CA: Multivariate Software, Inc.
- [5] Billón, M., Ezcurra, R., & Lera-López, F. (2008). The spatial distribution of the internet in the European Union. Does geographical proximity matter? *European Planning Studies*, 16(1), 119-142.
- [6] Billón, M., Ezcurra, R., & Lera-López, F. (2009). Spatial effects in website adoption by firms in European regions. *Growth and Change*, 40(1), 54-84.
- [7] Bou, J.C., & Satorra, A. (2007). The persistence of abnormal returns at industry and firm levels: evidence from Spain. *Strategic Management Journal*, 28, 707-722.
- [8] Bou, J.C. & Satorra, A. (2014). "Univariate versus multivariate modeling of panel data". WP 417, Universitat Pompeu Fabra, Department of Economics and Business.
- [9] Chaudhuri, A., Flamm, S., & Horrigan, J. (2005). An analysis of the determinants of internet access. *Telecommunications Policy*, 29(9-10), 731-755.
- [10] Çilan, C. A., Bolat, B. A., & Coşkun, E. (2009). Analyzing digital divide within and between member and candidate countries of European Union. *Government Information Quarterly*, 26(1), 98-105.
- [11] Colecchia, A. (2000). Defining and measuring electronic commerce - towards the development of an OECD methodology. Paris: OECD Press.
- [12] 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.
- [13] Cruz-Jesús, F., Oliveira, T. & Bacao, F. (2012). Digital divide across the European Union. *Information & Management*, 49, 278-291.
- [14] 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.
- [15] Dutta, S., & Jain, A. (2004). The networked readiness index 2003-2004: Overview and analysis framework. In S. Dutta, B. Lanvin & F. Paul (Eds.), *The global information technology report 2003-2004* (pp. 3-22). New York: Oxford University Press.
- [16] Grubestic, T. H. (2006). A spatial taxonomy of broadband regions in the United States. *Information Economics and Policy*, 18(4), 423-448.
- [17] Hanafizadeh, M. R., Saghaei, A., & Hanafizadeh, P. (2009). An index for cross-country analysis of ICT infrastructure and access. *Telecommunications Policy*, 33(7), 385-405.
- [18] Holloway, D. (2005). The digital divide in Sydney. A socio-spatial analysis. *Information, Communication & Society*, 8, 168-193.
- [19] Horrigan, J. B., Stolp, E., & Wilson, R. H. (2006). Broadband utilization in space: Effects of population and economic structure. *The Information Society*, 22, 341-356.

- [20] Howard, P. N., Anderson, K., Bush, L., & Nafus, D. (2009). Sizing-up information societies: Towards a better metric for the cultures of ICT adoption. *The information society*, 25, 208-219.
- [21] ITU. (2009). Measuring the information society. The ICT development index 2009. Geneva: ITU.
- [22] Kiiski, S., & Pohjola, M. (2002). Cross-country diffusion of the Internet. *Information Economics and Policy*, 14, 297-310.
- [23] Kolko, J. (2000). The death of cities? The death of distance? Evidence from the geography of commercial internet usage. In I. Vogelsang, & B. M. Compaine (Eds.), *The internet upheaval* (pp. 73-97). Cambridge, MA: MIT Press.
- [24] Mills, B. F., & Whitacre, B. E. (2003). Understanding the non-metropolitan-metropolitan digital divide. *Growth and Change*, 24(2), 219-243.
- [25] Noce, A. A., & McKeown, L. (2008). A new benchmark for internet use: A logistic modeling of factors influencing internet use in Canada, 2005. *Government Information Quarterly*, 25(3), 462-476.
- [26] OECD, D. L. (2001). Understanding the digital divide. Paris: Organization for Economic Cooperation and Development.
- [27] Orbicom. (2003). Monitoring the digital divide...and Beyond.(G. Sciades, Ed.) Retrieved from <http://orbicom.uqam.ca/upload/files/research_projects/2003_dd_pdf_en.pdf>.
- [28] Progressive Policy Institute. (2002). The 2002 state new economy index: Benchmarking economic transformation in the states. Retrieved from <<http://www.itif.org/files/2002-new-state-econ-index.pdf>>.
- [29] Satorra, A., & Bentler, P. M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations, *Computational Statistics & Data Analysis*, 10, (p. 285-305) (A. van Eye and C.C. Clogg, Editors). Sage Publications.
- [30] Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In *Latent Variable Analysis in Developmental Research* (p. 285-305).
- [31] Satorra, A. and P.M. Bentler (2014), "A Longitudinal Model for Repeated Cross Sectional Data with Clustering" Proceedings of the JSM 2013 - Business and Economic Statistics Section, pp. 3297-3311, American Statistical Association, Alexandria, VA 22314-1943.
- [32] Schleife, K. (2010). What really matters: Regional versus individual determinants of the digital divide in Germany. *Research Policy*, 39(1), 173-185.
- [33] Selhofer, H., & Hiising, T. (2002). The digital divide index: A measure of social inequalities in the adoption of ICT, in Wrycza, S. (Ed.), *Proceedings of the Xth European Conference on Information Systems ECIS 2002 - Information Systems and the Future of the Digital Economy*, 1273-1286. June 6-8, 2002. Gdansk: ECIS.
- [34] Soupizet, J. F. (2004). The north-south digital divide Paper presented at the digital divides international conference, November 18-19, Université Paris Sud, Paris.
- [35] US Department of Commerce, T. (2000). *Falling through the net: Toward digital inclusion*. Washington, DC: US Government Printing Office.
- [36] Vehovar, V., Sicherl, P., Hüsing, T., & Dolnicar, V. (2006). Methodological challenges of digital. *The Information Society*, 22(5), 279-290.
- [37] Vicente, M. R., & López, A. J. (2006). A multivariate framework for the analysis of the digital divide: Evidence for the European Union-15. *Information & Management*, 43, 756--766.
- [38] Vicente, M. R., & López, A. J. (2011). Assessing the regional digital divide across the European union-27. *Telecommunications Policy*, 35(3), 220-237.
- [39] Vicente, M. R., & López, A. J. (2007). Measuring the digital divide at the regional level: Evidence for Spain. Paper presented at the regional studies association

international conference: Regions in Focus?, April 2-5, University of Lisbon, Portugal.

- [40] Wolcott, P., Press, L., McHenry, W., Goodman, S., & Foster, W. (2001). A framework for assessing the global diffusion of the internet. *Journal of the Association for Information Systems*, 2(1).

Appendix A: Names of the areas

The following table exhibits the complete names of the 41 administrative units of Catalonia and the corresponding abbreviations used in this paper.

Table A.1: Complete names of the areas and their abbreviations

Name of area	Abbreviation	Name of area	Abbreviation
Alt Camp	ACA	Montsià	MON
Alt Empordà	AEM	Noguera	NOG
Alt Penedès	APE	Osona	OSO
Alt Urgell	AUR	Pallars Jussà	PJU
Alta Ribagorça	ARI	Pallars Sobirà	PSO
Anoia	ANO	Pla d'Urgell	PLU
Bages	BAG	Pla de l'Estany	PLE
Baix Camp	BXC	Priorat	PRI
Baix Ebre	BXE	Ribera d'Ebre	RIE
Baix Empordà	BXM	Ripollès	RIP
Baix Llobregat	BXL	Segarra	SGA
Baix Penedès	BXP	Segrià	SGR
Barcelonès	BAR	Selva	SEL
Berguedà	BER	Solsonès	SOL
Cerdanya	CER	Tarragonès	TAR
Conca de Barberà	CBA	Terra Alta	TAL
Garraf	GRF	Urgell	URG
Garrigues	GRG	Vall d'Aran	VAR
Garrotxa	GRT	Vallès Occidental	VOC
Gironès	GIR	Vallès Oriental	VOR
Maresme	MAR		

Appendix B: Socioeconomic characteristics of the areas

The following table reports the (Pearson) correlations between the available socioeconomic variables and the across-time average ICT activity of the areas in the CFA model.

Table B.1: Correlation matrix of socio-economic characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	1.000											
(2)	0.256	1.000										
(3)	0.486	0.258	1.000									
(4)	0.773	-0.005	0.086	1.000								
(5)	-0.824	-0.173	-0.353	-0.836	1.000							
(6)	0.262	0.184	0.455	-0.110	-0.004	1.000						
(7)	0.285	0.083	0.098	0.280	-0.242	0.054	1.000					
(8)	0.285	0.083	0.099	0.279	-0.242	0.054	1.000	1.000				
(9)	0.285	0.083	0.100	0.278	-0.242	0.054	1.000	1.000	1.000			
(10)	-0.613	-0.371	-0.381	-0.432	0.587	-0.218	-0.206	-0.206	-0.206	1.000		
(11)	-0.486	-0.608	-0.210	-0.340	0.336	-0.311	-0.218	-0.217	-0.217	0.398	1.000	
(12)	0.353	-0.214	0.607	0.125	-0.438	0.359	0.020	0.022	0.023	-0.332	0.173	1.000
(1)	Average of ICT activity (CFA)						(7) Employment in industry, 2007 (% of population)					
(2)	GDP per capita, 2008						(8) Employment in construction, 2007 (% of population)					
(3)	College degree, 2001 (% of population)						(9) Employment in services, 2007 (% of population)					
(4)	Population < 14 years old, 2008 (% of population)						(10) GVA in agriculture, 2008 (% of total)					
(5)	Population > 65 years old, 2008 (% of population)						(11) GVA in construction, 2008 (% of total)					
(6)	Population density, 2008						(12) GVA in services, 2008 (% of total)					

Correlations are based on 41 observations

Source: Idescat

The correlation between the GDP per capita and the average ICT activity of the areas (administrative units) is lower than 0.3. Other authors found a correlation of approximately 0.9 at the cross-country level (Hanafizadeh, Saghaei & Hanafizadeh, 2009) and approximately 0.7 at the regional level (Vicente & López, 2011). The correlation between the employment share in agriculture and the average ICT activity was approximately zero, and we do not report it in table B.1. For the rest of the main economic sectors, the correlations with the average ICT activity are lower than 0.3.

The differences in the productive structures are most likely better measured by the shares of the gross value added in agriculture, industry, construction and services. Here, the correlation between the share of the gross value added in agriculture and the average ICT activity is -0.613. The correlations between the shares in construction or services and the average ICT activity are both less than 0.5 in absolute value. The correlation for industry was approximately zero and has not been reported in the table.

Human capital is measured by the percentage of the population with a college degree in 2001, the latest observation available in our data. The correlation between this percentage and the average ICT activity is approximately 0.5, very close to the value found by Vicente & López (2011) at the cross-regional level. Population density does not seem to have any relevance to the ICT level of the areas. In contrast, the shares of the population under 14 and over 65 years of age are strongly correlated to the factor scores.

Appendix C: Estimation results for the panel data model

The following table displays the estimated factor scores of the panel data model. The areas have been ordered according to the value of the permanent component (last column).

Table C.1: Estimates of the factor scores of the panel data model

Area	f_1	f_2	f_3	f_4	q_1	q_2	q_3	q_4	p
	2008	2009	2010	2011	2008	2009	2010	2011	
MAR	0,934	1,085	1,150	1,171	0,053	0,203	0,269	0,290	0,881
VOC	0,876	0,893	1,038	1,062	0,040	0,056	0,202	0,226	0,836
TAR	0,864	0,965	1,079	1,015	0,051	0,152	0,267	0,202	0,813
BXC	0,798	0,836	0,930	0,964	0,041	0,080	0,173	0,208	0,756
BXM	0,827	0,928	0,939	0,745	0,097	0,199	0,209	0,016	0,730
BXL	0,799	0,881	0,809	0,796	0,071	0,153	0,081	0,068	0,728
GIR	0,762	0,845	0,846	0,815	0,050	0,133	0,134	0,103	0,712
ANO	0,784	0,783	0,709	0,697	0,082	0,081	0,007	-0,005	0,703
GRF	0,699	0,881	0,854	0,821	0,036	0,218	0,191	0,158	0,663
VOR	0,635	0,880	0,809	0,784	0,011	0,257	0,186	0,161	0,623
PLE	0,615	0,750	0,712	0,613	0,042	0,177	0,140	0,041	0,573
VAR	0,571	0,603	0,547	0,709	0,053	0,085	0,029	0,191	0,518
OSO	0,503	0,568	0,425	0,526	0,042	0,107	-0,037	0,064	0,462
APE	0,395	0,405	0,510	0,513	0,000	0,010	0,115	0,118	0,395
ACA	0,393	0,503	0,451	0,479	0,021	0,130	0,078	0,106	0,372
BAR	0,372	0,523	0,685	0,615	0,009	0,160	0,322	0,252	0,363
BAG	0,226	0,242	0,237	0,267	0,012	0,028	0,023	0,053	0,214
SEL	0,113	0,089	0,130	-0,093	0,023	-0,002	0,040	-0,184	0,090
PLU	0,094	-0,007	-0,061	0,068	0,043	-0,058	-0,112	0,017	0,051
GRT	0,046	0,041	0,066	0,050	0,005	-0,001	0,025	0,008	0,041
URG	0,017	0,058	0,004	-0,087	0,018	0,059	0,005	-0,087	0,000
SGR	-0,078	-0,167	-0,153	0,034	-0,014	-0,103	-0,089	0,098	-0,064
ARI	-0,101	-0,050	0,056	0,018	-0,015	0,035	0,142	0,104	-0,086
SGA	-0,155	-0,197	-0,033	-0,083	-0,046	-0,088	0,076	0,026	-0,109
AEM	-0,158	-0,238	-0,269	-0,272	-0,005	-0,085	-0,116	-0,119	-0,153
BXP	-0,225	-0,281	-0,464	-0,478	-0,012	-0,068	-0,251	-0,265	-0,213
CER	-0,344	-0,442	-0,321	-0,320	-0,050	-0,148	-0,027	-0,026	-0,294
PSO	-0,407	-0,360	-0,379	-0,352	-0,041	0,006	-0,013	0,014	-0,366
BXE	-0,373	-0,410	-0,461	-0,537	-0,007	-0,043	-0,094	-0,171	-0,367
SOL	-0,456	-0,405	-0,407	-0,467	-0,056	-0,005	-0,006	-0,067	-0,400
NOG	-0,410	-0,431	-0,506	-0,520	0,010	-0,011	-0,086	-0,100	-0,420
CBA	-0,494	-0,710	-0,641	-0,608	-0,033	-0,250	-0,181	-0,148	-0,460
PJU	-0,447	-0,644	-0,797	-0,737	0,037	-0,160	-0,313	-0,253	-0,484
RIE	-0,535	-0,595	-0,608	-0,579	-0,038	-0,097	-0,111	-0,082	-0,497
AUR	-0,697	-0,683	-0,565	-0,545	-0,088	-0,074	0,044	0,063	-0,609
BER	-0,707	-0,809	-0,686	-0,571	-0,079	-0,182	-0,059	0,056	-0,627
MON	-0,674	-0,755	-0,812	-0,853	-0,042	-0,123	-0,179	-0,221	-0,632
GRG	-0,930	-1,063	-1,066	-0,985	-0,034	-0,167	-0,170	-0,089	-0,896
RIP	-0,986	-1,171	-1,127	-1,080	-0,070	-0,255	-0,212	-0,164	-0,916
PRI	-1,310	-1,274	-1,561	-1,619	-0,075	-0,039	-0,326	-0,383	-1,236
TAL	-1,836	-2,064	-2,068	-1,974	-0,142	-0,370	-0,374	-0,280	-1,694

Appendix D: Standardized solutions as extracted from EQS output

The standardized solutions for equations (2), (3) and (4) are displayed below.

$$f_{1i} = 0.907 p_i + 0.422 q_{1i}$$

$$f_{2i} = 0.815 p_i + 0.580 q_{2i}$$

$$f_{3i} = 0.804 p_i + 0.597 q_{3i}$$

$$f_{4i} = 0.802 p_i + 0.422 q_{4i}$$

$$q_{2i} = 0.242 q_{1i} + 0.970 v_{2i} \quad R^2 = 0.059$$

$$q_{2i} = 0.357 q_{1i} + 0.934 v_{2i} \quad R^2 = 0.127$$

$$q_{2i} = 0.369 q_{1i} + 0.930 v_{2i} \quad R^2 = 0.136$$

$$p_i = 0.683 x_{1i} - 0.367 x_{2i} + 0.427 \delta_i \quad R^2 = 0.818$$