Micro data analysis of internet use in Europe

Albrecht Wirthmann, Eurostat

Abstract

The paper presents the results of a first attempt to explore the micro data of the community survey on the use of information and communication technologies in households and by individuals at the level of the European Union. The frequency of internet use and of downloading of audiovisual content are analysed using logistic regression models in order to isolate the influence of different explanatory variables. Socio-economic background characteristics such as age, gender, educational attainment, employment situation, geographic location, household income and composition, as well as the type of connection or place of internet access are used to explain differences in frequency of internet use and downloading. The advantage of using logistic regressions is that the influence of one explanatory variable can be quantified while controlling for the other variables,

The analysis shows that most of these explanatory variables influence the propensity for intensively using the internet and downloading audiovisual content. A very strong influence on the daily use of the internet has the educational attainment. The higher the education level the more likely the person uses the internet daily. Other very important factors are internet access from home and accessing the internet with different devices. The propensity for intensive downloading of films and music is very much influenced by the age, the employment situation, the gender and the place of internet access.

Introduction

Information and communication technologies (ICT) are having an increasing influence on the daily life of citizens. In particular, the internet is creating a new layer of participation of individuals in the societal and economic life. New services on the internet affect a wide range of activities such as information retrieval, communication, training and education, commerce and finance, participation in social networks or leisure activities. Activities that have been performed offline are now offered as online services on the internet. The growing pervasiveness of ICT leads to user friendlier applications on the one hand but requires access to ICT and internet and computer skills on the other hand. Persons who do not participate in the digital world run a risk to be excluded from important aspects of social end economic life.

The i2010 strategy[1] was designed as the strategic framework for the information society and media policies of the European Union. It promotes an open and competitive digital economy and emphasises the roles of ICT as driver for inclusion and quality of life. Related to this strategy a benchmarking framework has been elaborated that sets out a list of indicators to monitor the development of the information society in Europe. In addition to these annually collected core indicators, statistical data on a special topic are collected as special module. These data provide a deeper insight in certain aspects of the information society, i.e. use of advanced services, e-skills, security and trust or e-commerce. Based on a European regulation, the Member States of the EU collect harmonised statistics on the use of ICT in households and by individuals since 2006.

ICT for Inclusion is one of the main policy areas identified in the third pillar of i2010 initiative. Inclusion is defined as empowerment and participation of every individual in Europe in the information society, irrespective of age, gender, socio-economic status or ethnic background. Aggregated statistics and

[1] See http://ec.europa.eu/information_society/eeurope/i2010/strategy/index_en.htm

national studies on statistical data suggest that socio-economic characteristics affect the way in which the individual is exposed to and benefits from ICT.

Recognising the importance of e-Inclusion the 2006 Ministerial "Riga Declaration" on ICT for an inclusive society[2] has defined digitally disadvantaged groups set concrete targets for internet usage and availability, digital literacy and accessibility of ICT by 2010.

Purpose of the analysis

The aim of the study is to explore and analyse links between access to and the use of ICT and socio-economic characteristics of users and non-users of ICT. The study analyses different aspects of e-inclusion based on availability of statistical data coming from the European statistical survey on the access to and the use of ICTs by households and individuals. The analysis is based on micro-data, which are available for the majority of the Member States of the European Union at Eurostat.

In addition, the study serves as a feasibility study to assess the potential of the micro data for creating additional statistics on the issue of participation in the information society and e-Inclusion.

In the context of the study analysis in the following topic is analysed:

- Analysis and modelling of the intensity of internet use and activities as function of socio-economic background characteristics;

This study uses logistic regression models that predict the probability of the occurrence of an event on the basis of predictor variables, e.g. it is possible to predict the probability of a person for being an intensive internet user as a function of his age, gender or educational level. The occurrence of the event is modelled using the logistic function. The advantage of using the logistic regression is that the influence of the independent variables can be estimated variable by variable while controlling for the influence of the other independent variables.

The European survey on the use of information and communication technologies in households and by individuals

The study is derived from data of the annual survey on the use of ICT in households and by individuals in Europe. The main purpose of the statistical data collection is to provide information on the development of the European information society and thus to fulfil the needs for monitoring the various political initiatives at European and at national levels. The European survey is based on an annual Commission Regulation that determines the subjects, their characteristics, the coverage, the reference periods and the socio-economic background characteristics of the statistical data collection.

The survey consists of two parts. The first part collects information by household and the second focuses on individuals living in the household. The household part provides data on households' ICT equipment (devices, internet connection, broadband, etc). The second part contains questions on the individuals' frequency and location of computer and internet use, the purpose and nature of their activities on the internet and use of on-line services (e.g. for e-shopping, interaction with public services and administrations, e-learning, downloading content, arranging travel etc), e-skills and barriers to internet or broadband access. In addition to a defined set of core indicators, additional data is collected annually on a specific topic. The topics of the special modules are listed in the i2010 benchmarking framework[3]. In 2008, the special module focussed on advanced services[4].

---

[2] Riga ministerial conference "ICT for an Inclusive Society" (11-13.6.2006)
[3] See: http://ec.europa.eu/information_society/eeurope/i2010/benchmarking/index_en.htm
[4] See: http://circa.europa.eu/Public/irc/dsis/emisannexes/library?l=/data_-_database/theme_3_-_popul/isoc/methodological_informati&vm=detailed&sb=Title

The scope of the survey is limited to households with at least one member aged between 16 and 74 years and individuals within this age range. In order to be able to analyse differences in access and use of ICT, the survey additionally collects a number of socio-economic background variables including age, gender, education attainment, employment situation, occupation, geographical location, type of locality, household composition, and household income. These background characteristics are mainly used for the purpose of analysing the digital divide in the context of the European eInclusion policy. In addition to these socio-economic variables, additional breakdowns are derived from filter questions, e.g. having a broadband connection or being a frequent internet user. The reference period for most of the questions is the first three months of the year. The period is kep stable to exclude seasonal effects. Questions on e-commerce and eGovernment usage refer to the year before the survey.

The survey is mandatory in the EU Member States and additionally conducted in countries of the European Economic Area and accession or candidate countries to the EU. In 2008, the total net sample size was about 159 000 households and 232 000 individuals within the European Union. Almost all surveys are using face-to-face or telephone interviews. All participating countries transmit aggregate data to Eurostat following a defined transmission format. Starting from 2007, some countries provide Eurostat with individual data records, which do not allow direct identification of the respondents. In 2008, a total of 23 countries transmitted micro data to Eurostat on a voluntary basis. The micro data offers new potential for statistical data analysis. This paper presents a new approach of data analysis, which is not possible with the current tabulated data. However, the analysis suffers from the lack of micro data from a number of big Member States, i.e. Germany, France, the United Kingdom and Poland. Hence it is not possible to draw conclusions on the level of the European Union. The situation will change in the survey year 2011, when transmission of micro data will become mandatory for the Member States of the European Union.

Preparation of the database

The analysis of the micro data comprises data from 21 countries, 19 Member States and Iceland and Norway. Due to data validation problems at the time of the data analysis, the data from Czech Republic and Malta could not be included. Table 1 gives an overview on the data availability.

Table 1: Micro data availability, for the survey data on ICT use in households and by individuals, 2008

Country		Micro data		Households		Individuals	
Name	Code	available	included in analysis	Population in 1000	Share in analysis (%)	Population in 1000	Share in analysis (%)
Bulgaria	BG	Yes	Yes	2,723	3.4%	5,968	3.6%
Belgium	BE	Yes	Yes	4,054	5.1%	7,806	4.7%
Czech Republic	CZ	Yes	No	3,917		8,132	
Denmark	DK	Yes	Yes	2,258	2.8%	4,012	2.4%
Germany	DE	No	No	35,133		63,143	
Estonia	EE	Yes	Yes	553	0.7%	1,026	0.6%
Greece	EL	Yes	Yes	3,701	4.7%	8,236	5.0%
Spain	ES	Yes	Yes	15,080	19.0%	34,498	20.7%
France	FR	No	No	23,369		43,396	
Ireland	IE	Yes	Yes	1,466	1.8%	3,254	2.0%
Italy	IT	Yes	Yes	21,101	26.6%	44,885	27.0%
Cyprus	CY	Yes	Yes	248	0.3%	576	0.3%
Latvia	LV	Yes	Yes	844	1.1%	1,778	1.1%
Lithuania	LT	No	No	1,388		2,571	
Luxembourg	LU	Yes	Yes	156	0.2%	358	0.2%
Hungary	HU	Yes	Yes	3,445	4.3%	7,618	4.6%

Malta	MT	Yes	No	132		311	
Netherlands	NL	Yes	Yes	6,542	8.2%	12,053	7.2%
Austria	AT	Yes	Yes	3,216	4.0%	6,243	3.8%
Poland	PL	No	No	12,592		29,056	
Portugal	PT	Yes	Yes	3,503	4.4%	8,024	4.8%
Romania	RO	No	No	6,818		16,947	
Slovakia	SK	Yes	Yes	1,897	2.4%	4,190	2.5%
Slovenia	SI	Yes	Yes	649	0.8%	1,588	1.0%
Finland	FI	Yes	Yes	2,300	2.9%	3,877	2.3%
Sweden	SE	Yes	Yes	3,746	4.7%	6,745	4.1%
United Kingdom	UK	No	No	22,618		43,683	
Total	**EU27**	**21**	**19**	**183,452**	**43.3%**	**369,976**	**45.0%**
Iceland	IS	Yes	Yes	110	0.1%	212	0.1%
Norway	NO	Yes	Yes	1,827	2.3%	3,411	2.1%
Croatia	HR	No	No	1,452		3,427	
Total		**2**	**2**	**3,389**		**7,050**	

It has to be noted that 45.6% of the households and 47.7% of the individuals represented in the analysis are from Spain and Italy. The aggregate of the variables are therefore very much determined by these two Member States.

Background variables

The statistical analysis uses the available socio-economic background variables as explanatory variables for modelling the intensity of internet use, downloading contents from the internet or for performing different activities online. According to the division of the survey into a households' part and into an individuals' part, the background variables are related to either of them. Joining both parts enables attaching the household variables to the individual.

Table 2: Socio economic background variables

Variable name	Code	Variable type	Description
AGECLS		Individual	Age classes
	2		16-24
	3		25-34
	4		**35-44 (Reference group)**
	5		45-54
	6		55-64
	7		65-74
SEX		Individual	Sex
	1		**Male (Reference group)**
	2		Female
ISCED		Individual	Educational level
	0		**Primary or lower secondary education, no formal education ISCED 0-2, (Reference group)**
	3		Upper secondary education, ISCED 3,4
	5		Tertiary education, ISCED 5,6
EMPST		Individual	Employment situation
	1		**Employee or self-employed (incl. family workers) (Reference group)**
	2		Unemployed
	3		Student (not in the labour force)
	4		Other not in the labour force (retired, inactive, in compulsory military service, etc.)
GEO_DENS		Household	Type of locality
	1		Densely-populated area

Variable name	Code	Variable type	Description
	2 **3**		Intermediate area **Thinly-populated area (Reference group)**
HH_IQ	 1 **2** 3 4	Household	Household income quartile Lowest quartile **Second lowest quartile (Reference group)** Second highest quartile Highest quartile
HH_A_C	 **10** 11 20 21 30 31	Household	Household composition **Single household** One adult with one or more children Two adults Two adults with one or more children Three or more adults Three or more adults with one or more children
HH_CHILD	 **NO** YES	Household	Household composition **Without children** With children

The variable household composition has been included into the models in two variations. The first series of variables distinguishes between the number of adults living in the household and the presence of children. The second binary variable only makes a distinction concerning the presence of children in the household. The latter variable should provide information on the influence of the presence of children concerning access to and use of ICT. The first set of variables should be able to make further distinction in relation to patterns of access and use of ICTs by size of household.

In addition to the above background characteristics, additional variables were introduced as explanatory variables in the analysis. These are "having a broadband connection", "having used the internet at home within the last 3 months", "having paid for audiovisual content" and 4 additional variables describing wireless or mobile access to the internet. These additional variables could add to explaining the intensity of internet use and the characteristics of the different types of internet users.

Table 3: Additional explanatory background variables

Variable name	Code	Variable type	Description
BB	 **0** 1	Household	Do you use a broadband connection ('DSL'=1 or 'BBOTH'=1)? **No (Reference group)** Yes
IHM	 **0** 1	Individual	Did you use the internet at home in the last 3 months? **No (Reference group)** Yes
AVPAY	 **0** 1	Individual	Did you pay in the last 3 months for online audiovisual content? **No (Reference group)** Yes
IUMPH	 **0** 1	Individual	Do you access the internet with a mobile phone via GPRS? **No (Reference group)** Yes
IU3G	 **0** 1	Individual	Do you access the internet with a mobile phone via UMTS (3G)? **No (Reference group)** Yes
IUPALM	 **0** 1	Individual	Do you access the internet with a handheld computer (palmtop, PDA) **No (Reference group)** Yes

Variable name	Code	Variable type	Description
IUPORT		Individual	Do you access the internet with a portable computer (laptop) via wireless connection away from home or work
	0		**No (Reference group)**
	1		Yes

The values in the tables above which are marked in bold have been taken as reference for the logistic regression analysis, which is described below.

Modelling intensity of internet use as a function of socioeconomic background variables

The study focuses on the analysis of the internet use and tries to explore factors that determine or are related to its intensity of use.

Table 4: Percentage of individuals who used the internet within a defined time period by socio-economic breakdown within EU27, 2008

		Within last 3 months (Internet users)	Within last year	More than 1 year ago	Never used the internet
Gender	female	59	62	38	35
	male	64	67	32	30
Household income	lowest	40	43	56	53
	2nd	47	50	50	47
	3rd	63	66	34	32
	highest	78	80	19	18
Education level	low	40	42	57	55
	medium	67	71	29	26
	high	89	91	9	8
Broadband connection		87	89	11	10
Type of locality	urban	67	70	30	27
	intermediate	62	65	35	33
	rural	52	55	44	42
Employment situation	inactive	29	31	68	65
	active	74	77	23	21
	students	94	96	4	3
	unemployed	52	57	43	38
Age group	16 - 24	88	91	9	7
	25 - 34	78	81	18	16
	35 - 44	72	75	25	22
	45 - 54	60	63	37	35
	55 - 64	42	45	55	52
	65 - 74	20	21	78	76
TOTAL		62	64	35	33

Table 4 shows statistics on the last use of the internet broken down by socio-economic background characteristics. Persons who have used the internet within the last three month before the survey are classified as internet users.

62% of the individuals aged between 16 and 74 within EU27 are internet users. The difference between the internet users and those who had used it within the last year is quite small. Most persons who use the internet seem to be regular users, i.e. having used the internet within the last months.

Comparable to the access to the internet in households there is a difference in the last use of the internet by socio-economic characteristic. A higher percentage of men than women use the internet. The share of internet users increases with the household income. The percentage of internet users within the highest income quartile is almost double than the percentage in the lowest income quartile. Persons living in urban areas are more likely to be internet users than persons living in rural areas. Students and economically active persons show higher percentages of internet users than inactive or unemployed persons. The share of individuals who have used the internet within the last 3 months declines sharply by age group.

In addition to the last access to the internet, the survey collects statistics on the frequency of internet use. Table 5 and Table 6 display statistics on the frequency of internet use for the European Union (EU27) and the European aggregate (EU19+2), which is generated for the purpose of the micro data analysis.

Table 5 Frequency of internet use by background characteristics in EU27 in % of individuals, 2008

Socio-economic background characteristics		Internet use			
		Intensive (daily)	Weekly	Daily or weekly	Less than weekly or never
Age group	16 - 24	66	17	83	17
	25 - 34	57	14	72	28
	35 - 44	49	16	64	36
	45 - 54	39	14	53	47
	55 - 64	28	10	38	62
	65 - 74	11	6	17	83
Gender	male	47	12	60	40
	female	39	14	53	47
Education level	l0_2	25	10	35	65
	l3_4	45	15	61	39
	l5_6	70	14	85	15
Employment situation	inactive	17	8	25	75
	active	52	16	68	32
	students	76	15	91	9
	unemployed	33	12	45	55
Type of locality	urban	49	13	62	38
	intermediate	43	14	56	44
	rural	33	14	46	54
Household income	lowest	26	9	35	65
	2nd	29	12	41	59
	3rd	42	14	56	44
	highest	59	14	73	27
Broadband connection		66	16	82	18
Sum		**43**	**13**	**56**	**44**

Table 6: Frequency of internet use by background characteristics in EU19+2 in % of individuals, 2008

Socio-economic background characteristics	Daily	Weekly	Daily or weekly	Less than weekly or never	Total by category
AGECLS					

16 – 24	67	14	81	19	14
25 – 34	56	12	69	31	19
35 – 44	47	12	59	41	21
45 – 54	38	11	50	50	19
55 – 64	26	8	34	66	16
65 – 74	9	4	13	87	12
SEX					
Male	46	10	57	43	50
Female	38	11	48	52	50
ISCED					
l0_2	21	8	28	72	43
l3_4	51	13	64	36	36
l5_6	71	13	84	16	21
EMPST					
Active	52	13	65	35	57
unemployed	30	9	39	61	6
Students	76	12	89	11	8
Inactive	14	6	20	80	29
GEO_DENS					
Urban	47	10	57	43	42
intermediate	40	9	49	51	26
Rural	37	12	49	51	31
HH_IQ					
1st	24	7	31	69	15
2nd	30	10	41	59	24
3rd	45	11	56	44	25
4th	60	10	70	30	27
HH_CHILDREN					
No	39	10	49	51	68
Yes	47	13	60	40	32
Broadband					
Yes	66	14	80	20	50
accessed internet at home within last 3 months					
No	50	27	77	23	9
Yes	78	17	94	6	48
Not applicable					43
Sum	**42**	**11**	**52**	**48**	**100**

The summary statistics of the aggregates, EU27 and EU19+2, seem to be comparable. The differences are around 2 – 3 percentage points for most of the figures, with a maximum difference of 6 percentage points. More important are the relations of the variables by background characteristics. In both tables we can observe the same behaviour concerning the background characteristics. In brief, there is a clear distinction between the different age groups concerning daily and occasional (less than weekly or never) internet use. The percentage of daily internet user decreases with higher age group whereas occasional or non-use increases by age group. There is a difference between the genders, a higher percentage of men use the internet on a daily basis, whereas more women are occasional or non-internet users. Only 21% of the individuals with lower educational attainment are daily internet users in EU19+2 and 71% of the individuals with higher education are intensive internet users. The percentages of occasional internet users by educational attainment are diametrical. More than ¾ of the students and half of the active population are daily internet users, whereas only 30% of the unemployed and 14% of the economically inactive populations use the internet daily. Almost half of the individuals living in urban areas use the internet daily while this percentage is only 37% for individuals in rural areas. Only 24% of individuals of a household of the lowest income quartile are daily internet users, but 60% of the individuals of households of the highest income quartile. Nearly

half of the individuals living in households with children are daily internet users and 39% of individuals living in a household without children use the internet daily. 2/3 of the individuals who are member of a household with broadband access are daily internet users and only 20% of these individuals are less frequent internet users. Almost 4/5 of the individuals who accessed the internet at home within the last 3 months are daily internet users. Nevertheless, half of the individuals who did not access internet at home during the last 3 months are daily internet users. This group makes up 4.5% of the total population. In summary the socio-economic and individual characteristics age, gender, educational attainment, the employment situation, the household income and the type of locality seem to have an influence on the frequency of internet use. However, it is not possible to determine and separate from the aggregate statistics the level influence of the different characteristics on the behaviour of persons as regards to frequency of internet use. Moreover, individuals might adhere to different socio-economic groups that diminish or enlarge their propensity as regards to intensive or occasional use of the internet. Persons with a higher education might at the same time have a higher income and be economically active. Or they might use the internet frequently although they are unemployed or economically inactive. Students are mostly part of the younger age groups, have a higher educational attainment but might live in a household of the lowest income group.

The following statistical analysis tries to isolate these different types of influences by socio-economic and other background characteristics in order to measure their influence on the behaviour towards use of the internet independent from each other and controlling for their mutual influence.

Micro data analysis

The purpose of the analysis of the micro data is to explain in a statistical way different patterns of the intensity of internet use with the help of additional background information. In most cases socio-economic background information on the age of the individuals, the gender, the educational attainment the employment situation, the location of the household, the income of the household have been included in the analysis. In addition to these characteristics, we also introduced variables, such as "having a broadband connection" or "using the internet at home" as additional explanatory variables.

The analysis was done using a logistic regression, which predicts the probability of the occurrence of an event with the help of a number of predictor variables, the socio-economic background characteristics. The probability for the occurrence of an event is modelled using the logistic function. The dependent variable is dichotomous whereas the independent variables can be either continuous or categorical. For the purpose of the analysis, the background characteristics were converted into binary values, indicating the membership of an individual to one of the categorical values. This was done for the purpose of calculating odds ratio estimates that indicate the probability of an event compared to a reference group. Values > 1 indicate a higher probability, whereas values < 1 indicate a lower probability, e.g. individuals with higher education have a 2.4 times higher probability of being an daily internet user than individuals with lower education.

The logistic regression uses the logistic function as base formula. It takes as input any value between $-\infty$ and $+\infty$ whereas the result is limited to 0 and 1. In the case of the logistic regression, the output expresses the probability for the occurrence of a certain event. In order to determine the occurrence of an event (Y=1), it is assumed that an empirically not observable variable “Z” exist, which can produce the binary characteristic of the dependent variable Y as a function of the independent variables X_j.

Figure 1: Equation of the logistic regression:

$$p_k(y=1) = \frac{1}{1+e^{-z_k}}$$ p is the probability of success

with $z_k = \beta_0 + \sum_{j=1}^{J} \beta_j \cdot x_{jk} + u_k$ z are called the logits

Odds: $\frac{p(y=1)}{1-p(y=1)} = e^z$

The model of the logistic regression function assumes the existence of a non-linear relation between the probability of success of the binary variable P(y=1) and the independent variables. However, the combination of the independent variables of the aggregated factor Z is modelled as a linear combination. The probability for the occurrence of an event is calculated as relative probability in comparison to a certain reference group in the case of dichotomous explanatory variables.

The variable frequency of internet use was split into 3 dichotomous variables. Intensive internet users are those who use the internet daily or almost every day, frequent internet user access the internet weekly but not every day and occasional users are those who access the internet less than weekly including those who have ever used the internet. In the 2008 ICT use questionnaire, there has been a question on the frequency of downloads of music or films, which was answered by regular internet users, i.e. having accessed the internet within the last 3 months. The classification of the frequency to intensive, frequent and occasional downloaders was done in the same way as for the internet users. The model was calculated for the aggregate of the European countries as well as for the countries separately. Table 7 shows the odds ratios for the European aggregate, which consists of 19 EU Member States plus Iceland and Norway.

All 6 models provided significant results concerning the null hypothesis for the likelihood ratio, the Score and the Wald test could be reject with an error of < 0.001, i.e. at least one of the regression coefficients is not equal to zero. The tests for the individual predictors are in the majority highly significant except for certain variables that are indicated in the following table. Odds ratio with 3 stars are significant on the level of 99.9%, with 2 stars on 99% and with one star on 95% level. "C" is an indicator for the quality of the logistic regression and measures the area below the ROC curve. The ROC curve allows to compare at least two models and determines the one with the best fit. Possible values for C range from 0.5 corresponding to a model which predicts the result randomly to 1 corresponding to a model perfectly discriminating the response. Values between 0.7 and 0.8 represent an acceptable predictive ability and values > 0.8 represent excellent predictive quality. For most of the models C is close to 0.8, except for frequent internet users with only 0.628 indicating only a poor quality of the logistic regression model.

Table 7: Odds Ratio estimates of logistic regressions for intensity of internet use and downloading of audiovisual content in EU19+2, 2008

	Internet user			Downloading of audiovisual content		
Explanatory variables	**Intensive (daily)**	**Frequent (weekly)**	**Occasional**	**Intensive (daily)**	**Frequent (weekly)**	**Occasional**
AGECLS2	1.283 ***	0.934 *	0.696 ***	2.821 ***	3.305 ***	0.270 ***
AGECLS3	1.112 ***	0.887 ***	0.954	1.663 ***	1.730 ***	0.559 ***
AGECLS5	0.847 ***	1.093 ***	1.194 ***	0.447 ***	0.656 ***	1.746 ***

AGECLS6	0.858 ***	1.126 ***	1.134 ***	0.238 ***	0.432 ***	2.823 ***
AGECLS7	0.690 ***	1.273 ***	1.340 ***	0.213 ***	0.349 ***	3.418 ***
SEX2	0.763 ***	1.208 ***	1.238 ***	0.528 ***	0.614 ***	1.853 ***
ISCED3	1.546 ***	0.857 ***	0.596 ***	0.937	0.969	1.042
ISCED5	2.384 ***	0.712 ***	0.330 ***	0.707 ***	1.137 ***	1.020
EMPST2	0.815 ***	0.840 ***	1.676 ***	2.032 ***	1.064	0.697 ***
EMPST3	2.019 ***	0.613 ***	0.506 ***	1.670 ***	1.109 **	0.723 ***
EMPST4	0.577 ***	1.031	2.389 ***	1.138	1.108 *	0.866 ***
GEO_DENS1	1.342 ***	0.730 ***	0.941 **	1.582 ***	0.912 ***	0.918 ***
GEO_DENS2	1.290 ***	0.723 ***	1.029	1.714 ***	0.694 ***	1.078 **
HH_IQ1	*1.021*	*0.905* ***	*1.066*	*0.751* ***	*0.908* **	*1.202* ***
HH_IQ3	1.299 ***	0.787 ***	0.886 ***	0.942	0.854 ***	1.148 ***
HH_IQ4	1.685 ***	0.606 ***	0.774 ***	0.948	0.883 ***	1.122 ***
HH_A_C11	0.884	0.884	1.450 ***	0.544 ***	0.700 ***	1.574 ***
HH_A_C20	0.753 ***	1.164 ***	1.327 ***	0.819 ***	0.810 ***	1.265 ***
HH_A_C21	0.555 ***	1.323 ***	1.890 ***	0.636 ***	0.683 ***	1.571 ***
HH_A_C30	0.542 ***	1.383 ***	1.907 ***	1.019	0.805 ***	1.161 ***
HH_A_C31	0.446 ***	1.575 ***	2.292 ***	0.810 **	0.786 ***	1.316 ***
BB1	1.345 ***	0.750 ***	0.827 ***	1.320 ***	1.511 ***	0.655 ***
IHM1	5.152 ***	1.415 ***	0.074 ***	4.392 ***	2.997 ***	0.273 ***
AVPAY1	1.993 ***	0.661 ***	0.240 ***	0.835 ***	3.312 ***	0.371 ***
IUMPH1	1.577 ***	0.798 ***	0.483 ***	1.340 ***	1.528 ***	0.630 ***
IU3G1	1.566 ***	0.711 ***	0.631 ***	1.407 ***	1.098 **	0.797 ***
IUPALM1	2.160 ***	0.472 ***	0.568 ***	1.522 ***	0.872 **	0.917 *
IUPORT1	2.185 ***	0.582 ***	0.362 ***	1.417 ***	0.996	0.868 ***
C	**0.787**	**0.628**	**0.866**	**0.816**	**0.771**	**0.800**

Internet users

The influence of the age group on being an intensive, frequent or occasional internet user in "EU19+2" does exist. The odds ratios rank between 0.69 and 1.283 concerning the propensity for daily use of the internet and between 0.7 and 1.34 for occasional internet use. Individuals aged between 16 and 24 years have a 28% higher probability for being an intensive or daily internet user than individuals of the reference group between 35 and 44 years. Whereas individuals aged between 65 and 74 have a 45% lower probability for being and intensive internet user than the reference group. However, we can observe that the odds for individuals in age group 6 (55-64) are at the same level as for those in the younger age group 5 (45-54). The opposite results are calculated for the occasional or non-internet users with a 43% lower probability for the youngest age group and a 34% higher probability for the oldest age group to be an occasional internet user. The results support the observation from the aggregate statistics that the age of persons influences their propensity towards intensity of internet usage.

Women have a considerably lower probability of being an intensive internet user but a higher probability for frequent or occasional internet use compared to men. The educational attainment level

influences very much the probability of internet use intensity. Individuals with a tertiary education have a 2.4 times higher probability for being an intensive internet user and at the same time a 3 times lower probability for being an occasional internet user. The employment situation affects the probability of internet use intensity. Students are two times more likely to use the internet on a daily basis than individuals of the economically active population whereas the probability for the inactive population is almost two times lower than for the reference group. Unemployed and inactive persons have a considerably higher probability for being occasional internet users. Unemployed persons have a more than 22% lower probability for being intensive internet user and a 67% higher probability for being occasional or non-user than the economically active population. Persons living in intermediate and urban areas are more likely to be intensive internet users whereas they are less likely to use the internet on a weekly basis. Only a small difference can be observed or the figures are not significant for occasional internet users. Persons living in a household of the highest income quartile have a considerably higher chance to be intensive internet users. But the results for the lowest income quartile are not significant for daily and occasional internet use. Individuals who live in a household of the two lowest income groups have a similar propensity concerning frequency of internet use. This might due to the fact that the first quartile includes students, who have generally a rather strong propensity to use the internet daily.

Compared to the single person reference household type, the propensities of the other types of households for being a daily internet user are considerably lower. The odds decrease with the number of persons in the household. In addition, the presence of children decreases further the probability of the adults for using the internet daily. The opposite observation can be made for occasional or non-internet usage.

Accessing the internet via broadband influences the probability of being daily internet user positively. The strength is at the same level as for geographic location, each of them compared to the respective reference group. The variable with the highest influence as compared to the reference group is "using the internet at home". Users who access the internet at home are more than five times more likely to be a daily internet user than individuals who have not accessed the internet at home during the last 3 months. On the other hand the probability for being an occasional user when accessing the internet at home is more than 13 times lower than for individuals accessing the internet from elsewhere than home. These results suggest that the presence of internet access at home also leads to more intensive use of the internet. Accessing the internet from a PDA or a portable computer as well as the fact that users already have paid for audiovisual content increase the chances for being an intensive internet user considerably.

The logistic regression provides estimates for the odds of each explanatory variable while controlling for the other variables. The influence of the different explanatory variables can be combined and an estimate for being an intensive internet user with a specific combination of background characteristics as compared to the respective reference groups can be calculated.

Looking at the results of the logistic regression, being an intensive internet user is influenced by the age, the gender, the educational attainment level, the employment situation, the locality, the household income level, the household composition, the place of internet access, the fact that a person has already paid for audiovisual content and using additional devices to access the internet. Typically, a daily or intensive internet user is a man, has a high educational attainment and/or is a student, lives in intermediate or urban areas, lives in a single household with a high income, accesses the internet from home with a broadband connection and additionally accesses the internet with mobile devices. The probability for being an internet user for an individual with these characteristics is

more than 200 times higher than for individuals in the respective reference groups[5]. Whereas the typical occasional internet user is older, a woman, has a low educational attainment, is economically inactive or unemployed, belongs to a household with more than three members including children and does not have internet at home. The probability for those individuals is 7 times lower than for the individuals of the reference groups.

In addition to the analysis on EU19+2 aggregate level the logistic regression were performed country by country, too. However, it would lead too far to exhaustively describe the results of the logistic regressions by country. Therefore we have picked some observations that differ from the observation of the 'EU19+2'-aggregate. The size of the sample by country is considerably lower than for the aggregate. Therefore, the significances for the estimates tend to be lower. In general, the explanatory variables with the highest impact on the estimates show up with the highest significances and can be included in the discussion of the results.

The differences between the various household income quartiles are more distinct as compared to the EU19+2 aggregate, for example in Finland, Norway and Hungary. In Finland and Norway, individuals living in a household with the highest income quartile have a propensity of 2.3 and 2.7 to be an intensive internet user. The results for the other income quartiles are not significant. In Hungary, the probability for individuals in households of the highest income quartile is not so high (odds ratio 1.356) as that of the EU19+2 aggregate in relation to the respective reference group.

We can also note that in Belgium, just like in Norway and Finland, individuals belonging to the age class 16-24 years clearly have higher chances to be intensive internet users ($1.4 \leq$ odds ratio ≤ 4.6). The odds ratios decrease from younger age groups to the older age groups with the exception of the age groups 5 and 6 as already described for the EU19+2 aggregate. In Bulgaria, Estonia and Latvia, the odds for the 16 to 24 years old are higher as compared to the reference group (2.1 in BG – 3.8 in LV), while the odds for the oldest group are lower (0.33 in BG – 0.53 in LV). In Luxembourg, individuals in age class 5 (4454 years) have a higher probability for being intensive internet users than those of the reference group. In the Netherlands, individuals of the age class 35-44 years have a higher probability to be intensive internet users than person aged between 25 and 34 years. Young Portuguese individuals (age classes 16-24 and 25-34) have a much higher probability (odds ratio = 1.703) to be an intensive internet user as compared to the reference group at the 'EU19+2' level. In Sweden, the probability is 2.8 times higher (odds ratio = 2.826) for age class 16-24 years than for the reference group.

The educational level seems to have a very strong influence in Portugal. The odds ratio for individuals with a secondary education is 2.605 and individuals with a tertiary education even have a probability that is 4.824 times higher to be an intensive internet user than the reference group. This result corresponds to higher odds ratio for Portuguese students (odds ratio 3.83) compared to economically active individuals, which form the reference group.

In Italy, it is very interesting to note that the household income quartile, has a lower influence on the dependent variable as compared to the reference group than in the EU19+2 aggregate. The different household income quartiles have odds ratios varying from 0.856 (lowest income quartile) to 1.249 (highest income quartile) compared to the reference group. Individuals living in the lowest income quartile have lower chances to be intensive internet users. On the other hand, individuals from high income quartile households have higher chances but the probability is lower that the equivalent of the EU19+2 aggregate.

[5] The cumulative probabilitiy for the occurence of an event is multiplicative.

In Sweden it can be observed that the odds of unemployed persons for using the internet daily are higher than for persons who are economically active.

In Slovenia, the employment situation variable has the same behaviour as in the 'EU19+2'-aggregate, but with a stronger intensity, which means that Slovenian students have a much higher propensity to be an intensive internet users (odds ratio 5.443) as compared to the reference group, while unemployed (odds ratio 0.364), or economically inactive persons (odds ratio 0.211) have less chances than the economically active population.

In Estonia, women have a higher probability to be intensive internet uses than men (odds ratio 1.423).

In Latvia, individuals living in intermediate areas have a chance which is almost twice as high to be an intensive internet user than individuals living in rural areas. The estimates for individuals living in urban areas are not significant.

In Sweden, people living in densely-populated areas clearly have a higher propensity for being an intensive internet user than those living thinly-populated areas.

Norway and Finland reveal the biggest difference between the different types of regions of residence: individuals living in densely populated areas have a higher propensity for being intensive internet users than those ling in thinly-populated areas (odds ratio around 1.8 and 1.9).

Frequency of downloading films and music

The "C"-values of the logistic regressions for downloading audiovisual content indicate a good predictive ability of the chosen model. The results of the logistic regression for intensively (daily), frequently (weekly) and occasional (less than weekly) downloading audiovisual content from the internet follows by and large the pattern of using the internet. However, the odds ratios for certain explanatory variables are more extreme than for the frequency of internet usage and for some variables the odds ratios do not follow the pattern of internet usage frequency.

There is a more explicit graduation between the age groups concerning intensive and frequent downloading of audiovisual content. The odds ratios vary from 2.8 for daily and 3.3 for weekly downloading for the youngest age class. with a steady decrease for the 2 oldest age classes to less than 0.21 for daily and 0.35 for weekly downloading. The opposite pattern can be observed for occasional downloading. Younger individuals tend to download more often audiovisual content from the internet than older individuals, mirroring a age gap. Youngest generations are clearly in the information age and use internet also to access audio-visual content. The era of television as main audio-visual source is fading. This might heavily influence the distribution of audiovisual products in future. Women have only half the probability of men for being intensive or frequent downloaders.

In contrast to the internet use, downloading is not dependent on or even follows an opposite pattern as regards to educational attainment level. Individuals with tertiary educational level have less probability for being an intensive downloader. The estimates for secondary education level are not significant.

Unemployed persons have a two times higher probability for daily downloading audiovisual contents than economically active persons. There are only minor differences between the odds ratios for frequent downloaders by employment situation as compared to the reference group. However, the significance levels of the estimates are lower than for the other variables. Individuals living in urban and intermediate areas have a clearly higher probability for being an intensive downloader as compared to frequency of internet use. Individuals living in an household in the lowest income quartile are less likely to be daily downloaders of music or files. The propensity for weekly downloading is

highest for individuals living a household of the reference group, while persons of the reference households are less likely to be occasional downloaders.

Comparable to the internet use frequency persons living in single households have the highest probability for daily downloading audiovisual content. The presence of children diminishes this probability. Unlike for internet use frequency the propensity increases from a household with two to more persons.

Connecting to the internet via broadband increases the propensity for being frequent or intensive downloader and decreases the probability for being an occasional downloader. Like for internet use frequency, accessing the internet from home has the biggest influence on the propensity for intensive downloading and increases the probability by the factor 3 for frequent downloading. Obviously, downloading audiovisual content is an activity that is performed at home.

In contrast to internet use, the fact that individuals have paid for online visual contents diminishes slightly the propensity for being an intensive downloader of music and films (odds ratio 0.83). However, having paid for audiovisual contents increases the chances for frequent (weekly) downloading of audiovisual content by the factor of 3.3.

A typical person who intensively downloads films and music from the internet is young, male, unemployed or student, does not live in rural areas, lives in a one person household, has broadband connection and internet access at home. He is more inclined not to pay for downloading audiovisual content. This is in contrast to frequent downloaders who obviously are more willing to pay for downloading audiovisual content. The relationship between download and propensity to pay is therefore more complex than a simple negative one, showing various faces.

Frequent downloaders are young, male, live in rural areas in a single household, access the internet from home with broadband and have already paid for audiovisual content. Educational attainment and employment situation do not contribute to the probability for being a frequent downloader of films and music.

Conclusions

The study could prove the feasibility for performing micro data analysis on the data of the survey on ICT usage in households and by individuals. With the help of logistic regressions, the influence of a number of socio-economic background characteristics on the propensity of individuals concerning frequency of internet use and downloading of audiovisual content could be proved. The results of the study are preliminary as more countries should be included in the analysis. For the future, it is envisaged to intensify and to extend the analysis of the micro data to other areas of interest such as the digital divide. The micro data analysis enables users of the statistical data to isolate and quantify the influence of different parameters on the behaviour of individuals as regards to the access and use of ICT. This will provide valuable information to policy makers for the application of measures on policy level to achieve the goals in the various policy areas related to the information society.

References

Klaus Backhaus, Bernd Erichson, Wulff Plinke: Multivariate Analysemethoden, Berlin,2006

European Commisison: Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the

Regions - “i2010 – A European Information Society for growth and employment”, 2005, http://ec.europa.eu/information_society/eeurope/i2010/key_documents/index_en.htm

European Commission: Methodological manual for Statistics on the Information Society, 2008, http://ec.europa.eu/eurostat/ict

I2010 high level group: i2010 benchmarking framework, 2006, http://ec.europa.eu/information_society/eeurope/i2010/benchmarking/index_en.htm

Riga ministerial conference: "ICT for an Inclusive Society" (11-13.6.2006), http://ec.europa.eu/information_society/events/ict_riga_2006/index_en.htm