Internet Purchases by Individuals: Clusters of the Most vs the Least Developed European Countries

Ksenija Dumičić, Anita Čeh Časni, Berislav Žmuk
Department of Statistics
Faculty of Economics and Business, University of Zagreb
Zagreb, Croatia
kdumicic@efzg.hr, aceh@efzg.hr, bzmuk@efzg.hr

***Abstract* :**

The goal of the paper is to describe and analyse possible economic and ICT development level impacts on the Internet purchases by individuals, as the main variable under study, in selected European countries, EU28, and three official EU candidates, the Former Yugoslav Republic (FYR) of Macedonia, Serbia and Turkey. For this purpose, the Eurostat data for the main variable, as the dependent one, and for the additional four independent variables for 2013 are analysed.

Considering Internet purchases by individuals, the less developed analysed countries are mostly those being the South East European (SEE) countries. Four independent variables are selected, expressed as economic and Information and Communication Technology (ICT) development indicators and they are all shown to be positively correlated with the Internet purchases by individuals in selected European countries.

The best fitted multiple linear regression model developed shows that GDP per capita in Purchasing Power Standards (PPS), Index, EU-28=100, and Percentage of individuals using the Internet, are statistically significant in the regression model.

The main research hypothesis was that the SEE countries (SEECs) resemble each other concerning analysed variables, and they might belong to the same cluster. The k-means cluster analysis for 30 countries based on all five variables resulted with an acceptable four cluster solution. The clusters of the less developed, mostly SEE countries, and contain the EU official candidates: the FYR of Macedonia, Serbia and Turkey, but also the following three EU members: Bulgaria, Greece and Romania. The most developed EU countries are gathered in the cluster of their own with 10 countries: Austria, Belgium, France, Germany, Ireland, United Kingdom, and the Scandinavian countries: the Netherlands, Denmark, Finland and Sweden.

Keywords*—Internet purchases by individuals, ICT development indicators, Internet skills, Internet use, GDP per capita in PPS, multiple regression model, k-means clustering*

I. INTRODUCTION

The main variable under study here is "the Internet purchases by individuals, expressed as the percentage of individuals who purchased goods or services through Internet in the last 12 month", as it is called by Eurostat. In this paper the impact of new Information and Communication Technologies (ICT) use on purchasing is studied, assuming that higher level of Internet penetration and Internet literacy and skills together with an increased GDP per capita might influence increasing trends of Internet purchases by individuals.

After data exploration, this paper is studying impacts on percentage of individuals who purchased goods or services via Internet in the last 12 month using the correlation and linear regression analysis. The paper deals with the cluster analysis of countries, based on five variables especially focusing the clusters that gather the less developed and the most developed European countries. In that sense, the South East European countries (SEECs) are gathered are opposed the grouped most

developed European countries in the Northern Europe and Scandinavia.

E-commerce is defined as the sale or purchase of goods or services, whether between businesses, households, individuals or private organizations, through electronic transactions conducted via the internet or other online communication networks, see the Eurostat definition. The term covers the ordering of goods and services which are sent over computer networks, but the payment and the ultimate delivery of the goods or service may be conducted either on- or off-line. According to the Community survey on ICT usage in households and by individuals [1], e-commerce by individuals or households via the Internet is defined more specifically as "the placing of orders for goods or services via the internet", with e-mails orders excluded, and e-commerce is the most common form of distance shopping and has been growing steadily since it was first measured in 2004. Even 43% of EU consumers have purchased goods and services over the Internet in the 2011 year, which is 3 percentage points increase compared to 2010.

The recent Eurostat's Flash Eurobarometer [2] shows that shopping via the Internet continues to grow throughout the European Union (EU). More than 50% of European consumers did at least one online purchase in the twelve months before September 2012. This percentage has almost doubled compared to the year 2006. The Eurostat's report given in [3] shows many details about Internet use by individuals, households and enterprises. Eurostat figures show that e-commerce in Europe is booming, but that there is still much to be done before European e-commerce reaches its full potential. There are many opportunities, but also there are still many barriers that inhibit the growth of e-commerce, due to the differences in legislation, payments and logistic systems in Europe [4]. According to [5] socio-economic differences within a country are more important than differences between countries in explaining Internet access. Differences in access are explained by differing levels of GDP per capita. The mentioned authors find differences in individual characteristics to be less important than differences between countries and they give an insight to what extent Internet sales by individuals is influenced by the barriers people perceive to buying/ordering over the Internet, using of Internet, and level of computer/Internet skills they poses. Cluster analysis based on seven variables resulted with four clusters of similar countries. Reference [6] examines economic and ICT development indicators influence on recently increasing Internet purchases by individuals for European Union member states in 2011. Regression models developed showed that the GDP per capita and the ICT development level are essential for explaining the Internet purchases by individuals.

The research goal is to recognise the position of the less developed South-East European countries (SEECs) versus the rest of the Europe regarding the impacts of four development indicators taken on the main variable under study. The most developed North European, i.e. Scandinavian countries are focused, too. Data for 28 European Union countries (EU28), that include 6 SEE countries (Bulgaria, Croatia, Cyprus, Greece, Romania and Slovenia); and 3 SEE countries that are official EU candidates, the Former Yugoslav Republic of Macedonia (FYROM), Serbia and Turkey are used. Data for the rest four SEE countries, Albania and Montenegro, Kosovo and Bosnia and Herzegovina, are not available, so they could not be included into analysis.

II. DATA SOURCES AND DATA EXPLORATION

Variables impacting the dependent variable Internet purchases by individuals, $Y_{IntPurch}$, defined in Eurostat as "the percentage of all individuals who purchased goods or services through Internet in the last 12 months", included in this research for 2013 are: Gross Domestic Product per capita in Purchasing Power Standards (GDP per capita in PPS), Index, EU28=100, 2013; Public expenditure on education as percentage of GDP, data from 2010 are taken by the authors as the estimates for 2013 with the exceptions for Denmark

(estimate based on 2009), the FYR of Macedonia (estimate based on 2002), Greece (estimate based on 2005), Luxembourg (estimate based on 2001), Romania (estimate based on 2009, and Turkey (estimate based on 2006); Internet penetration rate (Internet use) given as Percentage of individuals using the Internet for 2013; and Individuals' level of Internet skills, percentage of the total number of individuals aged 16 to 74 for 2013, with the exceptions for the FYR of Macedonia (estimate based on 2010), and Serbia (estimate based on 2007) [7-9].

In 2013 the percentage of individuals who purchased via Internet in the last 12 months, $Y_{IntPurch}$, was in EU27 countries with data of 47% more than 135% higher than in the base year 2004 (20%). In EU15 countries (15 countries forming the EU before the enlargements of 2004 and 2007) the increase from 27% in 2004 to 53% in 2013, with the index of 196 (2004=100) shows a dramatic increase of Internet purchases of 96%, being nearly doubled, too. The analysed dynamics of the Internet purchases by individuals in the last 12 months, for the EU15 (15 countries forming the EU before the enlargements of 2004 and 2007) and for the EU27 countries, from 2004 to 2013, as given in Figure 2 and Figure 3, shows a yearly absolute increase of the linear trend with the slope of 3.1% for EU27 and 3.2% for EU15, being in both cases statistically significant at 5% significance level. Both linear trend models, calculated using ordinary least square estimators, are very representative with high coefficients of determination, explaining 99.2% of total variation in Internet purchases for the EU27, and 98.6% for the EU15 model.

When detecting the outliers in all variables included in the analysis, it should be noticed that only Luxembourg data for GDP per capita in PPS, with standardised value of higher then 4.1, should be excluded from the analysis, and that is why the data set is reduced from 31 to 30 countries.

The descriptive analysis for 31 countries is shown in Table 1. Based on 31 countries' data analysis for 2013, Table 1, the main variable under study, $Y_{IntPurch}$, with the range of 72, has the average of 38.81% and the coefficient of variation of 57.58%. The multiple Box-Plot displayed in Fig. 1 compares data variability for all 31 countries, showing outlier for two variables, a serious one for GDPpc in PPS (X_{GDPpc}) for Luxembourg, with Z_{LU}=4.1, and one mild outlier for Internet skills ($X_{IntSkill}$) for Germany, Z_{DE}=2.9, but it's data is kept for the further analysis.

The outlying Luxembourg data should be deleted from the analysis, so, the further calculations are made for 30 countries only.

Based on 30 countries' data analysis for 2013, after deleting Luxembourg data, the variable $Y_{IntPurch}$ remains with the same range of 72, and the average being slightly lower at of 37.77% and the coefficient of variation of 58.13%. There are 12 over averaged countries, and at the top there are: United Kingdom and Denmark (77%), followed by Sweden (73%), The Netherlands and Germany (69%), and Finland (65%). At the bottom, there are the SEECs: the FYR of Macedonia and Serbia (5%), Romania (8%), Turkey (10%), and Bulgaria (12%). Cyprus and Greece, both with 25% of individuals purchasing on-line in the last 12 months, are a little bit apart. The youngest EU country Croatia has the average 26%, and Slovenia 36%, which is close to the 30-countries average (37.77%), but a little bit away from the EU28 average for 2013 of 47%.

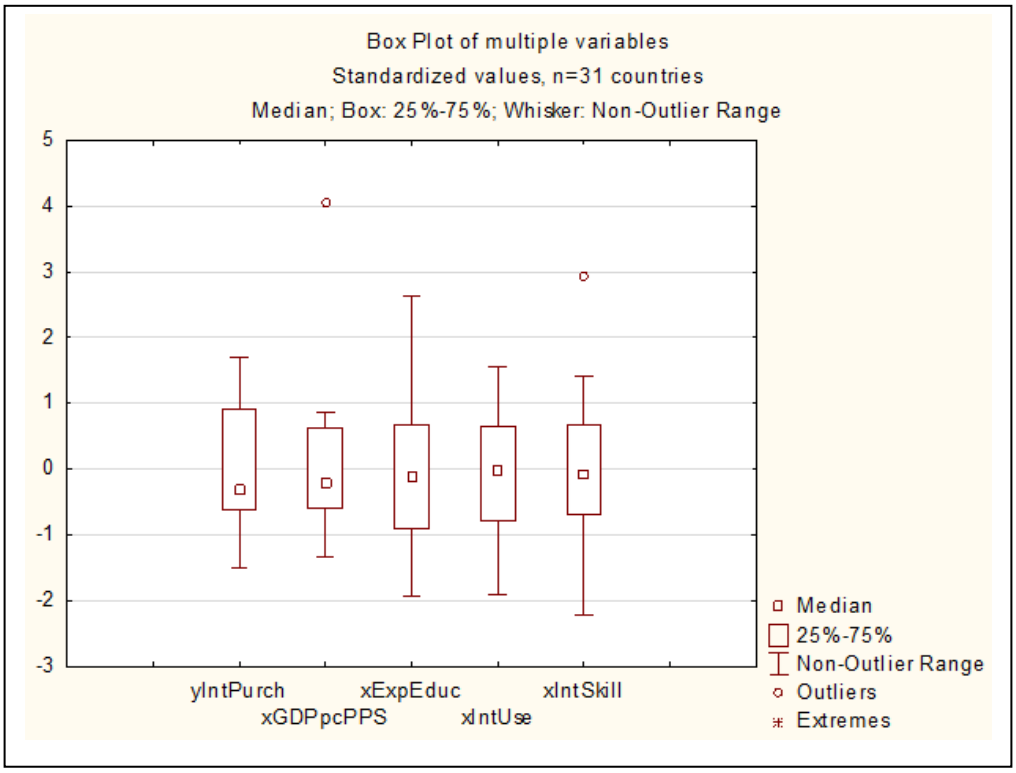
TABLE I. DESCRIPTIVE STATISTICS, N=31 COUNTRIES, 2013

Variable	Valid N	Mean	Median	Minimum	Maximum	Std. Dev.	Coef. Var.	Skewness	Kurtosis
$Y_{IntPurch}$	31	38.81	32.00	5.00	77.00	22.35	57.58	0.30	-1.03
X_{GDPpc}	31	92.19	83.00	35.00	264.00	42.49	46.08	2.17	8.07
$X_{ExpEduc}$	31	5.35	5.20	2.90	8.70	1.28	23.83	0.42	0.21
X_{IntUse}	31	73.04	72.70	46.30	94.80	14.00	19.16	-0.09	-0.83
$X_{IntSkill}$	31	26.61	26.00	12.00	46.00	6.59	24.75	0.64	1.46

Source: Eurostat, Authors' creation.

Fig 1. *The Box-Plot of standardized values for $Y_{IntPurch}$, X_{GDPpc}, $X_{ExpEduc}$, X_{IntUse} and $X_{IntSkill}$, in 2013 for n=31 countries (Eurostat, Authors' creation)*

*Note: Data for the variable $X_{ExpEduc}$ are taken as estimates for 2013 based on 2010



All the analysed variables for 30 countries have skewness coefficients close to zero, from -0.06 X_{IntUse} to 0.71 for $X_{IntSkill}$. The Anderson-Darling normality test was conducted for all the variables for 30 analysed countries in 2013, and it shows that the main variable under study $Y_{IntPurch}$ with the p-value=0.12, as well as other variables: X_{GDPpc} (p-value=0.16), $X_{ExpEduc}$ (p-value=0.82), X_{IntUse} (p-value=0.91), and $X_{IntSkill}$ (p-value=0.22), might be considered not to be significantly apart from the normal distribution at the 1% significance level.

III. CORRELATION AND REGRESSION ANALYSIS

The correlation coefficients for pairs of variables in correlation matrix show that the main variable under study, $Y_{IntPurch}$, called the dependent one, is positively correlated with each of the independent variables. The strongest positive correlation exists with X_{IntUse}, with r= 0.918, followed with the correlation with X_{GDPpc}, r=0.875. The variable $X_{ExpEduc}$ is with r=0.692 moderately correlated with $Y_{IntPurch}$. The weakest correlation of $Y_{IntPurch}$ exists with $X_{IntSkill}$, r=0.548.

All Possible Regression analysis with the dependent variable $Y_{IntPurch}$, and four (p=4) independent variables, X_{GDPpc}, $X_{ExpEduc}$, X_{IntUse} and $X_{IntSkill}$, was conducted. It resulted with altogether $m=2^p-1=2^4-1=15$ possible models. Because the strongest correlation between the dependent variable $Y_{IntPurch}$ arose to be with X_{IntUse} (r=0.918) and X_{GDPpc} (r=0.875), these two independent variables are put into the special focus when building the multiple regression model. Fortunately, these two

independent variables gave the multiple regression model with the best value for the Mallows' C_p index=3.513, compared to all other regression models which had C_p values between 5 and 154. So, it was decided to choose these two variables as the regressors for building the appropriate regression model which will be tested for diagnostics, too. The coefficient of determination of such a model equals R^2=0.8936 indicating that this model might be highly representative.

Substituting the calculated estimates and giving the main indicators of the respected multiple linear regression model using ordinary least squares estimators for the regression parameters, it follows:

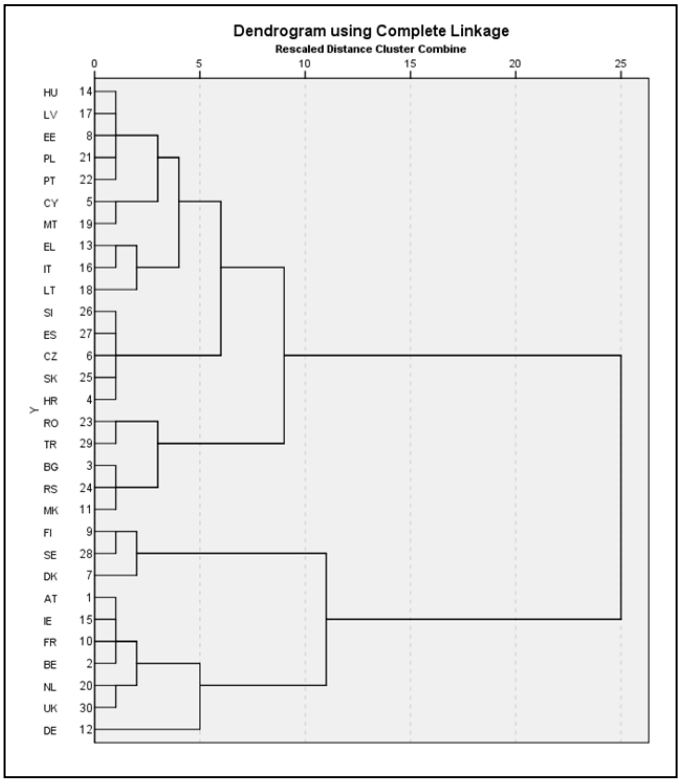
$$\hat{Y}_{IntPurch} = -58.2 + 0.3 \cdot X_{GDPpc} + 1.0 \cdot X_{IntUse} \quad (1)$$
$$(7.963) \quad (0.082) \quad (0.171)$$
$$n = 30,\ R^2 = 0.894,\ \hat{\sigma} = 7.421,\ \hat{V} = 19.65\%$$

According to (1), based on the overall F-test the whole model is statistically significant at 1% significance level (p-value 0.0000). The regressors X_{GDPpc}, with t-ratio=3.5834 and p-value 0.0013, and the regressor X_{IntUse}, with t-ratio=5.6862 and p-value 0.0000, are both statistically significant at the significance level of 1%. The regression model diagnostics results showed that there is no violation of the regression model assumptions. The coefficient of determination indicates that two regressors, X_{GDPpc} and X_{IntUse}, explain 89.39% of the total variation in $Y_{IntPurch}$. The coefficient of variation for the regression V-hat=19.65% indicates that the representativeness of the regression model is high. When interpreting the regression coefficients estimated for the multiple regression model developed for 30 countries in 2013, it should be said that the regression coefficient shows that if X_{GDPpc}, GDP per capita in Purchasing Power Standards (PPS), Index, EU28=100, would increase by one, without changing the level of the variable X_{IntUse}, the regression value of $Y_{IntPurch}$, percentage of all individuals who purchased goods or services through Internet in the last 12 months, would increase by 0.2945. The regression coefficient shows that if X_{IntUse}, Internet penetration rate (Internet use), given as Percentage of individuals using the Internet, would increase by one, without changing the level of the variable X_{GDPpc}, the regression value of percentage of all individuals who purchased goods or services through Internet in the last 12 months, $Y_{IntPurch}$, would increase by 0.9745.

IV. Hierarchical Cluster Analysis

After deleting the outlier, a seriously high value for X_{GDPpc} for Luxembourg, hierarchical clustering using standardized values of five analysed variables ($Y_{IntPurch}$, X_{GDPpc}, $X_{ExpEduc}$, X_{IntUse}, $X_{IntSkill}$) for 30 countries in 2013 was conducted. In the analysis complete linkage method was used. As measure of distances squared Euclidean distances were used. According to the dendrogram, what is given in fig. 2, the four cluster solution give the best interpretation possibilities. Also, after recognizing four clusters, further countries' division into more clusters results with too smaller differences, measured by cluster distances, between the clusters.

Fig. 2. Dendrogram of selected countries, used standardized values of variables $Y_{IntPurch}$, X_{GDPpc}, $X_{ExpEduc}$, X_{IntUse} and $X_{IntSkill}$, hierarchical clustering,



complete linkage method, squared Euclidean distances, data from 2013 for n=30 countries (Eurostat, Authors' creation)

*Note: Data for the variable XExpEduc are taken as estimates for 2013 based on 2010.

In the first cluster are 15 countries. There are mostly transition countries and that mostly recently become the EU member states like: Croatia, Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Portugal, Slovakia and Slovenia. But there are Greece, Italy and Spain as old EU15 member states. In the second cluster are placed less developed EU member states and the EU official candidate countries. This cluster has 5 following countries: Bulgaria, the FYR of Macedonia, Romania, Serbia and Turkey. The third cluster has 3 countries: Denmark, Finland and Sweden. Those countries could be observed as high developed Scandinavian EU state members. In the last, fourth, cluster are placed the most developed EU state members. This cluster counts 7 countries and they are: Austria, Belgium, France, Germany, Ireland, the Netherlands, and the United Kingdom.

V. K-MEANS CLUSTER ANALYSIS

After hierarchical, non-hierarchical, k-means clustering using standardized values of five analysed variables ($Y_{IntPurch}$, X_{GDPpc}, $X_{ExpEduc}$, X_{IntUse}, $X_{IntSkill}$) for 30 countries in 2013 was also conducted. The predefined four-cluster solution gathered the countries as it is shown in tables 2, 3, 4 and 5. Also, in the tables 2, 3, 4 and 5 there clusters with denoted Euclidean distances of each country's data around the cluster mean. The distances enable identifying potential cluster members which are very distant from the cluster centre. If this would be the case, the country apparently would not belong to the particular cluster. In all four clusters all distances from the respective cluster centre are considerably low and so it might be concluded that all countries have been placed in the appropriate clusters.

TABLE II. CLUSTER NUMBER 1: CLUSTER OF THE MOST DEVELOPED EU COUNTRIES

Country code for members of cluster number 1	**Distances from Respective Cluster Center Cluster contains 10 cases**
	Distance
AT	0.479932
BE	0.579148
DK	1.153423
FI	0.448667
FR	0.385518
DE	1.086276
IE	0.424493
NL	0.422060
SE	0.707826
UK	0.294035

The 1st cluster with 10 most developed EU countries includes: Austria, Belgium, Denmark, Finland, France, Germany, Ireland, the Netherlands, Sweden and the United Kingdom.

TABLE III. CLUSTER NUMBER 2: CLUSTER OF TRANSITION COUNTRIES THA MOSTLY BECOME THE EU MEMBER STATES IN THE RECENT YEARS

Country code for members of cluster number 2	**Distances from Respective Cluster Center Cluster contains 7 cases**
	Distance
HR	0.451014
CZ	0.422677
EE	0.409837
LV	0.327742
SK	0.277397
SI	0.343868
ES	0.369380

The 2nd cluster with 7 of countries that are mostly transition countries and that mostly recently become the EU member states includes: Croatia, Czech Republic, Estonia, Latvia, Slovakia, Slovenia and Spain. Spain is the only country from this cluster being an old EU member state.

TABLE IV. CLUSTER NUMBER 3: CLUSTER OF MOSTLY LOW TO MEDIUM DEVELOPED EU MEMBER STATES

Country code for members of cluster number 3	Distances from Respective Cluster Center Cluster contains 7 cases
	Distance
CY	0.600565
HU	0.418817
IT	0.581174
LT	0.611944
MT	0.627646
PL	0.345790
PT	0.142294

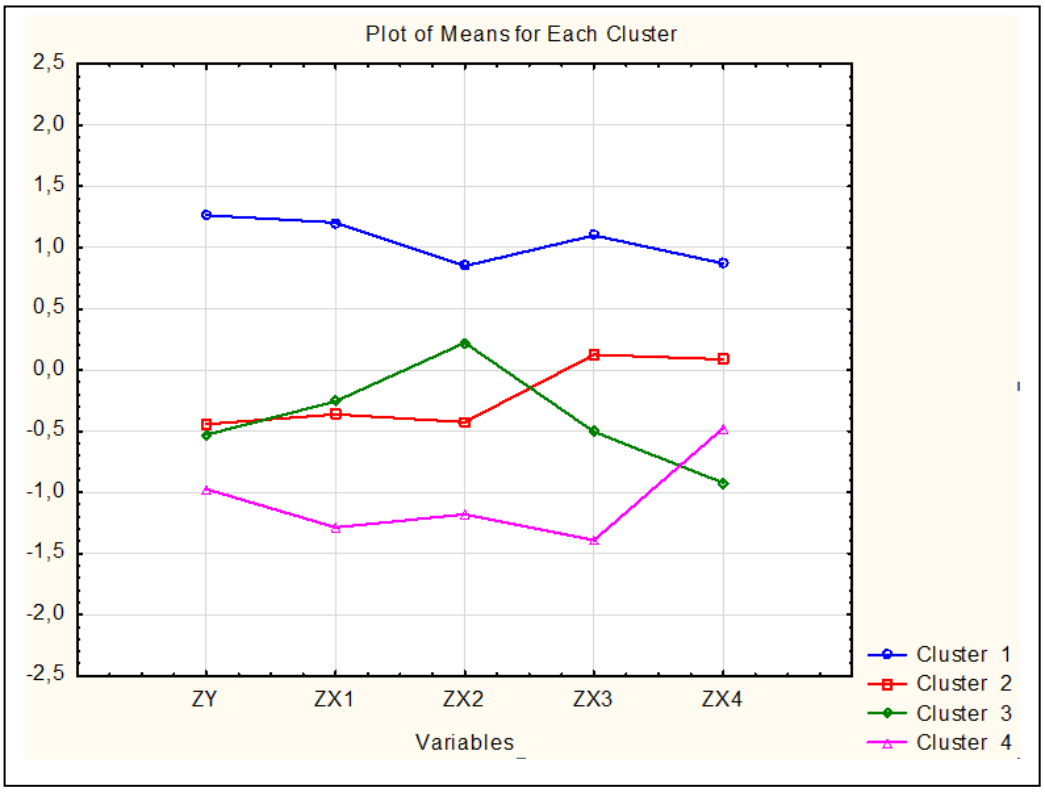
As given in table 4, the 3rd cluster covers mostly low to medium developed EU member states, such as Cyprus, Hungary, Italy, Lithuania, Malta, Poland and Portugal.

TABLE V. CLUSTER NUMBER 4: CLUSTER OF THE LESS DEVELOPED EU MEMBER STATES AND THE EU OFFICIAL CANDIDATE COUNTRIES

Country code for members of cluster number 4	Distances from Respective Cluster Center Cluster contains 6 cases
	Distance
BG	0.123631
MK	0.421791
EL	0.494641
RO	0.428121
RS	0.475908
TR	0.530819

The 4th cluster includes the less developed European countries, three of them being also the EU members. The EU official candidates the FYR of Macedonia, Serbia and Turkey belong to this cluster, but also the following three EU members: Bulgaria, Greece and Romania.

Fig. 3. *The Plot of standardized means for* $Y=Y_{IntPurch}$, $X1=X_{GDPpc}$, $X2=X_{ExpEduc}$, $X3=X_{IntUse}$ *and* $X4=X_{IntSkill}$, *in 2013 for n=30 countries (Eurostat, Authors' creation)*



*Note: Data for the variable $X_{ExpEduc}$ are taken as estimates for 2013 based on 2010.

The fig. 3 shows that the 1st cluster for the most developed countries has the highest values over all the analysed five variables. The 4th cluster collects the less developed countries, listed in table 5, has the lowest averages for all the variables, with an exception of the variable X4 for Internet Skills.

VI. Concluding Remarks

Economic and Information and Communication Technology (ICT) development results with an increase of Percentage of individuals who purchase by Internet in countries all over Europe, and so it is true for the less developed European studied here, such as the South-East European countries (SEECs).

The variable GDP per capita in PPS, Index, EU28=100, for 30 data in 2013 has the average 86.5, deviating relatively from the mean by 33%, showing moderate to large dispersion of data. 10 countries are above the EU28 average. The highest data for X_{GDPpc} in 2013 has Austria (129). Focusing the ranges, the less developed analysed countries are gathered at the bottom: the FYR of Macedonia (35), Serbia (36), Bulgaria (47), Romania (54), Turkey (55) and Croatia (61). Greece is in a little bit better position with data 75. The highest is the value for GDP per capita for Cyprus (86), even better then Slovenia who has GDP per capita of 83 (all in PPS, index, EU28=100). The arithmetic mean for the Internet penetration rate (X_{IntUse}) is 72.35% of people using the Internet, with relative deviation of 18.93%, and the range 48.53%, showing moderate to high dispersion of data. Countries with the lowest Internet penetration rates are in the SEE countries Turkey (46%), Romania (50%), Serbia (52%) and Bulgaria (53%). A little bit higher values are in Greece (60%), the FYR of Macedonia (61%), Cyprus (61%), Croatia (67%) and even (73%) in Slovenia.

All five analysed variables have great variability for 30 countries. The greatest variability appeared to be for the main variable under study, Percentage of individuals who purchase by Internet, with the coefficient of variation of 58%. The second highest variability with 33% is shown for X_{GDPpc}. The coefficient of variation for $X_{IntSkill}$, 25%, for $X_{ExpEduc}$ it is 23%, being moderately high. The less variable is the indicator X_{IntUse} with coefficient of variation of 18.93%. In the same time, all the analysed variables have distributions that are close to the normal, with skewness close to zero, from -0.06 X_{IntUse} to 0.71 for $X_{IntSkill}$.

Based on 30 countries' data analysis for 2013, the main variable under study, $Y_{IntPurch}$, with the huge range of 72% absolutely, has the average of around 38% and the coefficient of variation of a little bit above 58%, which shows that the variability of the percentage of the Internet purchases users over the analysed countries if quite large. At the top there are United Kingdom and Denmark with the data value of 77%. At the bottom, there are the FYR of Macedonia and Serbia (5%), Romania (8%), Turkey (10%), and Bulgaria (12%), all being the less developed European countries analysed in this paper. Cyprus and Greece, with data of 25%, are in a little bit better position regarding the main variable studied. Croatia (26%) and Slovenia (36%) are close to the 30-countries average of 37.8% in 2013.

When analysing correlations, the variable $Y_{IntPurch}$ is positively correlated with each of the independent variables. The strongest positive correlation exists with X_{IntUse}, with correlation of 0.918, followed with the correlation with X_{GDPpc}, r=0.875.

The hierarchical and non-hierarchical cluster analysis gave similar results. The k-means cluster analysis of 30 countries conducted here gave the four-clusters solution. The less developed analysed countries, those SEECs, gathered more or less together. As expected, the most developed EU countries are gathered in the cluster of their own.

The limitation of this research lays in the fact that Albania, Montenegro, Kosovo, and Bosnia and Herzegovina, being all part of the SEE region, are not included into the analysis because their data are not available.

REFERENCES

[1] Eurostat, "Community survey on ICT usage in households and by individuals," 2013, available at: http://epp.eurostat.ec.europa.eu/portal/page/portal/information_society/data/database (accessed: 10 June 2014).

[2] European Commission, "Consumer Attitudes towards Crossborder Trade and Consumer Protection," Flash Eurobarometer, vol. 358, 2013, available at: http://ec.europa.eu/public_opinion/flash/fl_358_en.pdf (accessed: 1 July 2014).

[3] European Commission, "Digital Agenda for Europe Scoreboard 2012," 2012, available at: http://ec.europa.eu/digital-agenda/sites/digital-agenda/files/KKAH12001ENN-PDFWEB_1.pdf. (accessed: 15 June 2014).

[4] Kommerskollegium, "E-commerce – New Opportunities, New Barriers, A survey of e-commerce barriers in countries outside the EU", 2012, available at: http://www.kommers.se/Documents/dokumentarkiv/publikationer/2012/skriftserien/report-e-commerce-new-opportunities-new-barriers.pdf (accessed: 1 August 2014).

[5] V. Šimičević, B. Jaković, and J. Ježovita, J., "Perceived Barriers to E-commerce: Empirical evidence from EU countries," Interdisciplinary Description of Complex Systems, vol. 11, 2013, pp. 123-130.

[6] K. Dumičić, A. Čeh Časni, and I. Palić, "Internet Purchases in European Union Countries: Multiple Linear Regression Approach," International Journal of Social, Human Science and Engineering, vol. 8, 2014, pp. 193-199.

[7] Eurostat, "GDP per capita in PPS", 2014, available at: http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&plugin=1&language=en&pcode=tec00114 (accessed: 1 June 2014).

[8] Eurostat, "Individuals' level of internet skills", 2014, available at: http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&plugin=1&language=en&pcode=tsdsc470 (accessed: 1 June 2014).

[9] Eurostat, "Internet purchases by individuals", 2014, available at: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_ec_ibuy&lang=en (accessed: 1 June 2014).

Technical Enablers for Cloud Computing Successful Adoption

Torki Altameem

Dept. of Computer Science, RCC, King Saud University,

P.O. Box: 28095 – 11437 Riyadh-Saudi Arabia.

altameem@ksu.edu.sa

Abstract :

The fast growth and development of the IT infrastructure in addition to the increase usage of the Internet globally has remarkably improved the capability to adopt cloud computing services. Using the cloud, it is easier, 24/7 accessible and better way to perform services than through traditional ways. Moreover, speed and simple access as well as lower costs for both clients and providers has also contributed in the growth of cloud computing. However, even with this growth, many have several concerns with cloud computing such as security. This research explores technical enablers for cloud computing adoption that can contribute to analyze organization's existing or future cloud initiative.

Keywords-component*; Teachnical enablers; cloud computing; Capacity; Virtualization*

I. INTRODUCTION

The fast growth and development of the IT infrastructure in addition to the increase usage of the Internet globally has remarkably improved the capability to adopt the cloud service. According to [1], cloud service providers propose many choices of services and applications to individuals, public and private sectors, for instance, information storing and sharing, instate usage of different software, database mining and management.

Cloud computing can be defined as "A standardized IT capability (services, software or infrastructure) delivered via Internet technologies in a pay-per-use, self-service way" [2]. In [3], Gupta stated that cloud computing services are easier, cheaper and better way to perform services than through traditional ways. Moreover, speed and simple access as well as lower costs for both clients and providers. In addition to pay for what you use model made growth and variety of offered services and develop the ability to integrate.

The concept of cloud computing and its growing popularity have gained many different businesses new opportunities and advantages. There are different types of cloud computing services. For example, there are Software as a service (SaaS), Infrastructure as a service (FaaS), Platform as a service (PaaS), Hardware as a Service (HaaS), and Data as a Service (DaaS) [4].

This concept offers a lot of services throughout the internet such as data storage, software, hardware and infrastructure. In addition, with cloud, there is no need for installing and running the application or software on the client's computer. Companies that offer cloud service provide their applications and services using Internet that are accessible anytime anywhere using desktop, tablets and mobile apps.

According to Allied Market Research Report that been published in July 2014, the cloud services in the world market is expected to

expand and reaching a market size of 555 billion Dollars in 2020. Furthermore, in 2014, the revenue of the global cloud services market will reach 209.9 billion Dollars [5].

However, even with this growth, many have several concerns with cloud computing. For example, unknown data storage location, loss control of the infrastructure and platform, integrity and availability of information, performance and future existence of the services provider company, maintain confidentiality and level of security at the service provider from attacks and vulnerabilities.

II. Technical Enablers for Cloud Computing

Studying literature review is performed by searching the academic publications and real cases form information technology, scientific computing and new technology adoption research fields. Our study suggests that these enablers influence the decision to adopt a cloud computing in an organization:

A. Virtualization

Virtualization [6] has enabled the abstraction of computing resources such that a single physical machine is able to function as a set of multiple logical Virtual Machines. Virtual Machines provide the ability to host multiple operating system environments which are completely isolated from one another on the same physical machine. Also, it offers the capability to configure Virtual Machines to utilize different partitions of resources on the same physical machine. Virtualization enables workloads to move between on-premises software and the cloud.

B. Capacity

It is referred to volume of work or data processing capacity that the cloud system can handle. It includes computing and storage capability. Many enterprises move to cloud computing without a clear and detailed capacity management strategy because cloud platform is seen as infinitely elastic, where capacity can be purchased as and when needed. But, there will always be limits to that elasticity though. Also, buying resources on the cloud instantly can be expensive and enterprises can relax that cost by planning for capacity in advance and avoiding over- or under-provisioning [7].

When IT executives upload smaller and familiar applications and workloads on to the cloud, they are aware of the capacity and manage it based on their expertise and experience.

C. Interoperability

It is referred to the ability of supporting heterogeneous cloud configurations. It includes applications to utilize multiple distributed heterogeneous resources. Cloud computing facilitates scalability and virtualized resources over internet as a service providing cost effective and scalable solution to customers. Cloud computing has evolved as a disruptive technology and picked up speed in 2008 and 2009 with the presence of many vendors in cloud computing space. With the presence of numerous vendors, the need is emerging for interoperability between clouds so that a complex and developed business application on clouds is interoperable [8].

D. Compatibility

It is universally acknowledged today that cloud computing works out to be the most cost efficient option for companies. However, the problem arises out of the fact that the company would have to replace much of its existing IT infrastructures in order to make the system compatible on the cloud.

Compatibility is another key that restrains large organizations from embracing cloud technology. When existing IT infrastructure may not be compatible with cloud technology or too complex to restructure, hybrid clouds might be the answer to data deployment. A third party provider will handle transition and reduce dramatically time and costs with personnel and technology for your business. One simple solution for this problem is to use the hybrid cloud, which is capable of addressing most of these compatibility issues [9].

E. Fault Tolerance

Since cloud computing environments are extremely heterogeneous and dynamic, with resources joining and leaving the cloud all the time, more faults are likely to occur in cloud environments. Also, the likelihood of errors occurrence is exacerbated by the fact that many cloud applications will perform long tasks that may require several days of computation. This will lead to a number of new conceptual and technical challenges to fault-tolerance researchers. The most important one is the scheduling of user jobs to cloud resources with meeting the user's Quality of Service (QoS) in existence of resource faults.

F. Security

Cloud computing infrastructure is a promising new technology and greatly accelerates the development of large scale data storage, processing and distribution. Private data of date owners are now placed on public clouds which are out of their trusted domains in cloud computing. Data owners do not have direct control over their sensitive data and are increasingly worrying about possible data loss and/or illegal use of their private data. Usually, cloud servers are considered as curious and untrusted entities. Data owners will hesitate to adopt cloud technologies if there are risks of data exposure to a third party or even the cloud service provider itself. Therefore, providing sufficient security protections on sensitive data is extremely important, especially for those applications dealing with health, financial and government data. To prevent information disclosure, the mainstream solution is to encrypt private data before uploading it onto the cloud server [10].

G. Trialability

Trialability is one of the most important components in the process of adopting a new technology. It represents a characteristic of cloud computing services and its potential has not been fully recognised as a determinant of adoption of cloud services.

"Trialability is defined as the degree to which an innovation may be experimented with on a limited basis" [1]. It refers to the ability to try or experiment with the performance of cloud computing services on a limited basis, with the benefits of characteristics of on demand, pay-per-use and try-before-you-buy. Trialability acts as a proxy to behavioural intent, or adoption.

H. Resiliency

Putting your IT assets into the cloud transforms a traditional server infrastructure into a dynamic environment that enables better utilisation of hardware and more flexible management of computing demand. However, a cloud environment is only as good as the resiliency built into the system. Whether you're looking to set up your own private cloud or employ the services of one of the many external cloud providers, it is essential that the cloud infrastructure is built and delivered with availability at its core if you are to stay in control of your critical systems [11].

I. Complexity

The cloud's complexity is a source of concern for some clients' employees, which can result in resistance to its adoption. Although training is of course one way to mitigate this concern, providers must also work to build systems that are intuitive to use[12].

III. CONCLUSION

This research explored technical enablers for cloud computing adoption that can contribute to analyze organization's existing or future cloud initiative.

Moreover, the research helps to deliver a better understanding of how these enablers impact on adoption cloud computing and in turn guide decision maker in organizations to a more knowledgeable decision regarding adoption of cloud computing. This research was exploratory, therefore, more work must be done carefully examine the implementation on the cloud computing and its technical enablers in order to achieve a better understanding on cloud computing adoption.

The researcher wants to perform several case studies on both cloud service providers and their clients. Using face-to-face interviews with senior decision makers in companies that provide the cloud services as well as interviewing their clients, the researcher hopes to identify the technical enablers for cloud computing adoption. This method will help the researcher to validate these technical enablers for better implementation of cloud computing.

References:

[1] E. Hand, Head in the clouds. New Jersey: Nature publication, 2007.

[2] J. Staten, T. Schadler, J. R. Rymer, and C. Wang (2009, August 14). Q&A: By 2011, CIOs Must Answer The Question, "Why Not Run In The Cloud?". (S. Leaver, & A. Viglianti, Interviewers)

[3] Anu. Gupta (2010), Cloud Computing Growing Interest and Related Concerns, 20IO 2nd International Conference on Computer Technology and Development (ICCTD 2010), pp.462-465.

[4] Sasikala P. and Chaturvedi Makhanlal : "Cloud Computing in Higher Education: Opportunities and Issues," International Journal of Cloud Applications and Computing, Vol. 1, Issue 2, April-June 2011, pp. 1-13.

[5] Allied Market Research. 2014. Global Cloud Services Market (Service Type, Cloud Type, Application and Geography) - Size, Share, Global Trends, Company Profiles, Demand, Insights, Analysis, Research, Report, Opportunities, Segmentation and Forecast, 2013 – 2020 [Online]. Available at: http://www.alliedmarketresearch.com/cloud-services-market.

[6] R. Buyya et al., “Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility,” Future Generation Computer Systems, Vol. 25, 2009, pp. 599-616.

[7] Mina Sedaghat, Francisco Hernandez, Erik Elmroth, “Unifying Cloud Management: Towards Overall Governance of Business Level Objectives,” In Proceedings of the 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2011), USA, 591-597, 2011.

[8] A. V. Parameswaran and A. Chaddha, "Cloud Interoperability and Standardization," Cloud Computing Infosys, vol. 7, 2009, pp. 19-26.

[9] A. V. Dastjerdi and R. Buyya, “Compatibility-Aware Cloud Service Composition under Fuzzy Preferences of Users,” IEEE Transactions on Cloud Computing, Vol. 2, No. 1, January-March 2014, pp 1-13.

[10] R. Li et al, ”Efficient multi-keyword ranked query over encrypted data in cloud computing,” Future Generation Computer Systems, Vol. 30, 2014, 179–190.

[11] http://www.sungardas.co.uk/Services/Cloud-Infrastructure/Cloud-Availability/Pages/Cloud-Availability.aspx.

[12] L. Morgan and K. Conboy, ”Key Factors Impacting Cloud Computing Adoption,” Computer, Vol. 46, Issue 10, 2013, pp. 97-99.