Editorial: Enterprise Websites Usage in the European Union - Outlook to the Western Balkans Countries

Ksenija Dumičić1, Berislav Žmuk2 and Mirjana Pejić Bach3
1University of Zagreb, Faculty of Economics and Business, Department of Statistics, Zagreb, Croatia, kdumicic@efzg.hr
2University of Zagreb, Faculty of Economics and Business, Department of Statistics, Zagreb, Croatia, bzmuk@efzg.hr
3University of Zagreb, Faculty of Economics and Business, Department of Informatics, Zagreb, Croatia, mpejic@efzg.hr

May 2016

Introduction

The Internet usage by individuals and enterprises has become unavoidable in the contemporary society [16], [20], [21], [25], [27], especially taking into account the European Commission’s Digital Agenda [6], [28]. Therefore, it is very important for enterprises to recognize usefulness and potential benefits of being present on the Internet [22], [27]. Although many enterprises use social media and Web 2.0 tools [2], [3], [23] there are still numerous enterprises that have not decided to open their own website yet [26].

There are many factors which influence the decision whether to have or not to have a website [25]. However, in this paper only variables related to the information technology (ICT) usage for e-commerce purposes (online ordering and buying), and availability of ICT infrastructure (e.g. broadband availability) were taken into account as possible determinants of the decision on the establishment of a company’s website. A similar approach was used by [1], [4], [15].

In the analysis, the ICT development levels (in terms of enterprise websites, e-commerce and ICT infrastructure) in selected European countries were observed, with a special focus on the following selected Western Balkans (WB) countries: Bulgaria, Croatia, Greece, Romania, Serbia, Slovenia, and the Former Yugoslav Republic of Macedonia (FYROM). Since Bulgaria, Croatia, Greece, Romania, and Slovenia are the EU member states, whereas Serbia and the Former Yugoslav Republic of Macedonia (FYROM) are not, it is expected that these five EU member states have a higher ICT development level than the two named non-EU member states, which is in line with the previous research [5], [19].

The paper is organized as follows. After a brief Introduction, Chapter 2 describes data and methods used in the research. In Chapter 3, results are discussed whereas in Chapter 4 conclusions are brought in full.

Data and Methods

The data used in the analysis are extracted from the Eurostat database. The list of all selected ICT variables is given in Table 1. The percentage of enterprises having a website or homepage is determined as the dependent variable of the research, while there are also two groups of potential independent variables: e-commerce-related variables and ICT infrastructure-related variables. E-commerce-related variables measure the percentage of enterprises having purchased via computer-mediated networks and having received orders via computer-mediated networks. ICT infrastructure-related variables measure variables such as the percentage of enterprises with a broadband access (fixed or mobile). All the information is presented as the percentage of enterprises, having at least 10 employees, in a particular country in 2015. Financial sector enterprises were not included in the analysis.

Data are collected for enterprises from 33 European countries, namely, for all the European Union member states (EU28), Iceland, Norway, Serbia, FYROM and Turkey. Even though the most recent data, from 2015, are collected for most of the countries, data from 2014 were used as an approximation for 2015 for Iceland, Serbia and FYROM. Similarly, because of the same reason data from 2014 were also used for the variable X Purch for Denmark and for the variable X Mobile/B2B for the United Kingdom. Unfortunately, the most recent data for the variable X Purch for Turkey are those for the year 2010, and this information was used as a proxy for the 2015 level, corrected by the increase in the variable X Order, which is available for the year 2015.
Table 1: ICT variables chosen for the analysis

<table>
<thead>
<tr>
<th>Variable code</th>
<th>Variable description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{Web}$</td>
<td>Enterprises having a website or homepage, All enterprises, without the financial sector (10 or more persons employed), Percentage of enterprises</td>
<td>[9]</td>
</tr>
<tr>
<td>$X_{Purch}$</td>
<td>Enterprises having purchased via computer-mediated networks, All enterprises, without the financial sector (10 or more persons employed), Percentage of enterprises</td>
<td>[7]</td>
</tr>
<tr>
<td>$X_{Order}$</td>
<td>Enterprises having received orders via computer-mediated networks, All enterprises, without the financial sector (10 or more persons employed), Percentage of enterprises</td>
<td>[12]</td>
</tr>
<tr>
<td>$X_{ComUse}$</td>
<td>Enterprises using computers, All enterprises, without the financial sector (10 or more persons employed), Percentage of enterprises</td>
<td>[8]</td>
</tr>
<tr>
<td>$X_{IntAcc}$</td>
<td>Enterprises with the Internet access, All enterprises, without the financial sector (10 or more persons employed), Percentage of enterprises,</td>
<td>[10]</td>
</tr>
<tr>
<td>$X_{BroAcc}$</td>
<td>Enterprises with a broadband access (fixed or mobile), All enterprises, without the financial sector (10 or more persons employed), Percentage of enterprises</td>
<td>[11]</td>
</tr>
<tr>
<td>$X_{FixBAcc}$</td>
<td>Enterprises using DSL or other fixed broadband connections (as of 2014), All enterprises, without the financial sector (10 or more persons employed), Percentage of enterprises</td>
<td>[11]</td>
</tr>
<tr>
<td>$X_{MobBAcc}$</td>
<td>Enterprises connecting to the Internet via a mobile broadband connection (3G modem or 3G handset), All enterprises, without the financial sector (10 or more persons employed), Percentage of enterprises</td>
<td>[11]</td>
</tr>
</tbody>
</table>

First, an outlier detection and analysis is performed due to the fact that outliers could have a significant impact on the results of the analysis. The outliers were omitted from the analysis. Second, after omitting detected outliers, descriptive statistical results are provided for each variable separately. The position of the observed WB countries according to the values of the observed ICT variables is additionally examined using profile diagrams. Third, the multiple linear regression analysis is used to determine which independent variables have a statistically significant impact on the changes in the dependent variable $Y_{Web}$. Fourth, the hierarchical cluster analysis was used to determine how the observed WB countries are positioned based on the following criteria: the occurrence of enterprise websites, the level of e-commerce development and the availability of ICT infrastructure in comparison to other observed European countries.

Results

In this part of the paper results of different statistical analyses of the observed ICT variables are provided. Following analyses are applied: outlier analysis, descriptive statistics analysis, multiple regression analysis, and hierarchical cluster analysis.

Outlier Analysis

In order to avoid and to reduce the possibility of obtaining misleading results by statistical analysis, first, an outlier analysis is conducted. However, before conducting an outlier analysis, all variable values are first standardized for each variable separately using the z-score approach. After that, the most commonly used technique in detecting outliers, the box-plot diagram, was constructed for each observed variable and it is shown in Figure 1.
The box-plots, which are given in Figure 1, reveal that there are outliers at two variables, $X_{\text{BroAcc}}$ and $X_{\text{FixBAcc}}$. At both variables Bulgaria was determined to be an outlier because in comparison to other European countries it has very low values of these two variables. To be more precise, at both variables Bulgaria has values which are more than three standard deviations below the average (Bulgaria’s z-score for the variable $X_{\text{BroAcc}}$ is −3.69 and for the variable $X_{\text{FixBAcc}}$, it is −3.46). Consequently, the data for Bulgaria have been omitted from further analysis. Accordingly, data for 32 countries are used in further analysis.

**Descriptive Statistics Analysis**

Descriptive statistics results for all the eight variables under the study are given in Table 2. All the observed variables are expressed in percentages. According to the coefficient of variation, very low data variability is present at variables $X_{\text{ComUse}}$ (CV=3.51%), $X_{\text{IntAcc}}$ (CV=3.82%), $X_{\text{BroAcc}}$ (CV=4.05%) and $X_{\text{FixBAcc}}$ (CV=5.24%). The main reason of very low data variability at these variables is the fact that certain countries have already reached or are very close to the maximum available at these variables (which is 100%). On the other hand, according to the coefficient of variation, the highest data variability is present at the variables $X_{\text{Purch}}$ (CV=48.48%) and $X_{\text{Order}}$ (CV=39.59%). Furthermore, only distributions of these two variables, $X_{\text{Purch}}$ and $X_{\text{Order}}$, are positively skewed whereas distributions of other variables are negatively skewed.

![Figure 1: Box-plots of standardized variables, data for n=33 countries, for 2015](image)

Table 2: Basic descriptive results, data for n=32 countries without Bulgaria, for 2015

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Y_{\text{Web}}$</td>
</tr>
<tr>
<td>Mean</td>
<td>75.31</td>
</tr>
<tr>
<td>Median</td>
<td>78.0</td>
</tr>
<tr>
<td>Mode</td>
<td>83</td>
</tr>
<tr>
<td>Variance</td>
<td>139.8</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>11.82</td>
</tr>
<tr>
<td>Coefficient of variation</td>
<td>15.70</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.60</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.04</td>
</tr>
<tr>
<td>Minimum</td>
<td>45.0</td>
</tr>
<tr>
<td>1st quartile</td>
<td>66.5</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>83.0</td>
</tr>
<tr>
<td>Maximum</td>
<td>95.0</td>
</tr>
<tr>
<td>Range</td>
<td>50.0</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>16.5</td>
</tr>
</tbody>
</table>

Within the sample of 32 European countries, there are seven countries which belong to the group of WB countries and which will be explored in more detail. For that purpose, profile diagrams with standardized data (z-scores) were used. The profile diagrams for Croatia, Greece, Romania and Slovenia, which are also the European Union member.
states, are given in Figure 2 whereas the profile diagrams for FYROM and Serbia, which are not the European Union member states, are given in Figure 3. Turkey, which is a Balkan country but is still not a member of the European Union, was also taken into account in this analysis.

According to Figure 2, Slovenia has in general a higher ICT development level than Croatia, Greece and Romania, measured in terms of enterprise websites, e-commerce and ICT infrastructure. However, Croatia, which is in the second placed among these four observed countries, seems to have a higher z-score at the variable $X_{MobBAcc}$ than Slovenia. It seems that there is a very low difference in the ICT development level between Greece and Romania. In both countries it is significantly under the average ICT development level of all the 32 observed European countries. On the other hand, Slovenia is slightly under the average only in the case of the variable $X_{Purch}$.

Figure 2: Profile diagrams for the observed Western Balkans countries – the European Union member states
Note: Line at the z-score 0 level indicates an average of the 32 observed European countries.

Figure 3: Profile diagrams for the observed Western Balkans countries – non-European Union member states
Note: Line at the z-score 0 level indicates an average of the 32 observed European countries.
Profile diagrams in Figure 3 show that Serbia is in general more advanced and has a higher ICT development level than FYROM. It has higher z-scores at all the variables except at the variable \( X_{\text{MobBAcc}} \). Furthermore, in comparison to all the 32 observed European countries, Serbia is under the average ICT development level only at the variables \( Y_{\text{Web}} \) and \( X_{\text{MobBAcc}} \). On the other side, both FYROM and Turkey seem to have a lower ICT development level in comparison to the average ICT development level of all the 32 observed European countries.

**Multiple Regression Analysis**

A multiple regression analysis was used to detect which of the independent variables, indicating the level of e-commerce development and availability of ICT infrastructure, have a statistically significant impact on the fact whether an enterprise has a website, as the main variable of this research. The main variable of interest is the Percentage of Enterprises having a website or homepage \( (Y_{\text{Web}}) \). Other seven ICT variables are used as independent variables.

Since the multiple regression model with all the seven independent variables did not meet all model requirements due to the multicollinearity of the independent variables, the number of independent variables in the model was reduced by using the backward stepwise regression approach [14]. The probability of F to entry was set to 0.05, and the probability of F to removal of a variable from the regression model was set to 0.10. The final regression model was reached in the step 4 when the next four independent variables were excluded from the regression model in the following order: \( X_{\text{FixBAcc}}, X_{\text{MobBAcc}}, X_{\text{Order}} \). In other words, variables \( X_{\text{Purch}}, X_{\text{FixBAcc}} \) and \( X_{\text{MobBAcc}} \) entered the final regression model as independent variables. In order to additionally test validity of the formed multiple regression model, the model diagnostics is conducted, and presented in the paragraph that follows.

Based on the overall F-test, the whole model was statistically significant at the 1% significance level \( (F\text{-ratio}=29.515, \ p\text{-value}<0.0000) \). Furthermore, all three regressors \( X_{\text{Purch}} \) \( (t\text{-ratio}=4.138, \ p\text{-value}=0.0003) \), \( X_{\text{FixBAcc}} \) \( (t\text{-ratio}=3.549, \ p\text{-value}=0.0014) \) and \( X_{\text{MobBAcc}} \) \( (t\text{-ratio}=3.641, \ p\text{-value}=0.0011) \) are statistically significant at the significance level of 1%. Because all Variance Inflation Factors are smaller than 5 \((X_{\text{FixBAcc}}=1.313, \ X_{\text{MobBAcc}}=1.098, \ X_{\text{MobBAcc}}=1.296)\), it has been concluded that the multicollinearity problem does not exist in the observed regression model. The Breusch-Pagan-Godfrey Heteroscedasticity test \( (\text{test statistics}=3.433, \ p\text{-value}=0.03295) \) and the White Heteroscedasticity test \( (\text{test statistics}=6.085, \ p\text{-value}=0.7314) \) show that there is no heteroscedasticity problem at the significance level of 5%. Similarly, the Durbin-Watson test \( (\text{test statistics}=2.234) \) and the Breusch-Godfrey Serial Correlation LM test \( (\text{test statistics}=0.93; \ p\text{-value}=0.6260) \) lead to the conclusion that the autocorrelation of the residuals problem does not exist at the significance level of 5%. Finally, the conducted Jarque-Bera Normality test \( (\text{test statistics}=0.290; \ p\text{-value}=0.8650) \) pointed out that non-normality of the residuals problem is not present in the regression model at the significance level of 5%. Accordingly, all the regression model assumptions are met, so the regression model estimates can be interpreted whereas the estimated regression model is given as:

\[
\hat{Y}_{\text{Web}} = -38.11 + 0.29 \cdot X_{\text{Purch}} + 0.84 \cdot X_{\text{FixBAcc}} + 0.37 \cdot X_{\text{MobBAcc}} \quad n = 32 \quad R^2 = 0.76 \quad \bar{R}^2 = 0.73
\]

\[
(21.38 \quad 0.07) \quad (0.24 \quad 0.10) \quad \bar{\sigma} = 6.10 \quad \bar{V} = 8.10%
\]

The adjusted coefficient of determination indicates that three regressors \( X_{\text{Purch}}, X_{\text{FixBAcc}} \) and \( X_{\text{MobBAcc}} \) explain 72.9% of the total variation at the dependent variable \( Y_{\text{Web}} \). The regression model can be considered as highly representative because the regression coefficient of variation is only 8.10%.

Therefore, the final regression model indicates that the variable that indicates the percentage of enterprises having purchased via computer-mediated networks \( (X_{\text{Purch}}) \) and the two variables which are indicators of the development of ICT infrastructure, representing the speed and the type of the Internet connection \( (X_{\text{FixBAcc}} \) and \( X_{\text{MobBAcc}} \) have a statistically significant impact on the main variable of interest - Percentage of Enterprises having a website or homepage \( (Y_{\text{Web}}) \).

**Hierarchical Cluster Analysis**

In order to research whether the selected WB countries which are the EU member states (Croatia, Greece, Romania, and Slovenia) have a higher ICT development level than the selected WB countries which are not the EU member states (Serbia and FYROM), a cluster analysis is conducted. In the analysis, the ICT development level is observed and measured through all the eight variables from Table 1.

Before any cluster analysis was conducted, all variables were standardized, and z-scores were used in the further analysis. Because the observed sample size is 32 countries, it has been decided to observe only four clusters solutions. This decision is based on the rule of thumb, according to which the number of clusters should be equal to the square root of the sample size divided by two [17]. The hierarchical cluster analysis is conducted based on squared Euclidean distances. In the cluster analysis, the complete linkage amalgamation (joining) rule is used [13].

Table 3 provides the number of countries in the formed clusters and the clusters’ mean z-scores. In order to calculate a mean z-score for countries in each of the four clusters, first, the average z-score for each country across
all the eight observed variables was calculated. After that a mean z-score of a cluster was calculated as an average z-score of all the countries in the observed cluster. Furthermore, the clusters are labelled with numbers from 1 to 4 where countries in the Cluster 1 have the highest average z-score or the highest ICT development level whereas countries in the Cluster 4 have the lowest.

Table 3: The number of countries in the formed clusters and the clusters' mean z-scores

<table>
<thead>
<tr>
<th>Amalgamation (joining) rule</th>
<th>Cluster 1: The highest ICT development</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4: The lowest ICT development</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of countries</td>
<td>Mean z-score</td>
<td>No. of countries</td>
<td>Mean z-score</td>
<td>No. of countries</td>
</tr>
<tr>
<td>Complete linkage</td>
<td>5</td>
<td>0.78</td>
<td>16</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Table 4 presents the distribution of countries based on the clusters for complete linkage hierarchical clustering approach with squared Euclidean distances taking into account data for 32 countries and 8 variables for 2015. The clustering results suggest that the observed WB countries can have a higher or a lower ICT development level regardless of whether they are the European Union member states or not. The same conclusion can be drawn from the profile diagrams of the observed WB countries which are given in Figure 2 and 3.

Table 4: Distribution of countries in the formed clusters

<table>
<thead>
<tr>
<th>Amalgamation (joining) rule</th>
<th>Cluster 1: The highest ICT development</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4: The lowest ICT development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete linkage</td>
<td>Denmark, Austria, Finland, Iceland, Norway</td>
<td>Belgium, the Czech Republic, Germany, Estonia, Ireland, Spain, France, Lithuania, Luxembourg, Malta, the Netherlands, Portugal, Slovenia, Sweden, the United Kingdom, Serbia</td>
<td>Italy, Cyprus, Latvia, Slovakia, FYROM</td>
<td>Greece, Croatia, Hungary, Poland, Romania, Turkey</td>
</tr>
</tbody>
</table>

Table 5 shows memberships to the clusters of the seven observed WB countries by using hierarchical clustering approach with squared Euclidean distances taking into account data for 32 countries and 8 variables for 2015. According to the clustering results it can be concluded that Slovenia and Serbia have the highest ICT development level among the observed WB countries. Croatia, FYROM and Turkey have a lower ICT development level than Slovenia and Serbia and are members of Cluster 3 and Cluster 4. On the other hand, Greece and Romania have convincingly the lowest ICT development level, not only among the observed WB countries, but also among all the observed European countries. Greece and Romania were classified into Cluster 4 which consists of countries with the lowest enterprise website occurrence, e-commerce development and availability of ICT infrastructure in comparison to other observed European countries.

Table 5: Membership to the same cluster of the seven observed WB countries

<table>
<thead>
<tr>
<th>Amalgamation (joining) rule</th>
<th>Country</th>
<th>European Union member states</th>
<th>Non-European Union member states</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete linkage</td>
<td>Croatia</td>
<td>Greece</td>
<td>Romania</td>
</tr>
</tbody>
</table>

Conclusions

The goal of the research was to explore the level of e-commerce development, the development of ICT infrastructure, and having an enterprise website among European countries using the Eurostat data for 2015. A special focus is given to selected WB countries, including Bulgaria, Croatia, Greece, Romania, Serbia, Slovenia, and the Former Yugoslav Republic of Macedonia (FYROM).

Our results indicate that Finland, Lithuania and the Netherlands achieved the highest development level concerning the level of e-commerce development, the development of ICT infrastructure, and having an enterprise website.
The multiple regression model revealed that the variable indicating the level of e-commerce development (percentage of enterprises having purchased via computer-mediated networks) and the two variables indicating the development of ICT infrastructure (the speed and the type of Internet connection) have a statistically significant impact on the decision of an enterprise to establish its own website. The other variable (indicating the percentage of enterprises having received orders via computer-mediated networks) did not have a statistically significant impact on the decision of an enterprise to establish its own website. These results indicate that companies are more eager to present themselves through the Internet to their suppliers and possibly other stakeholders than to their potential buyers. A possible reason for such behaviour is that enterprises are still faced with significant organizational and other barriers (e.g. such as trust) in using their own website as a selling channel [18].

Considering the Western Balkans countries, Slovenia has the highest ICT development level (measured by the abovementioned variables) in comparison to Croatia and FYROM. On the other hand, FYROM has the lowest ICT development level among the observed Western Balkans countries. Therefore, by using the hierarchical cluster analysis we have confirmed the results of the previous research [5], [19], indicating the presence of the digital divide in the European Union.

The main limitation of this study is that the remaining WB countries (Albania, Bosnia and Herzegovina, Montenegro and Kosovo) could not be included into the analysis since their data for the variables of interest are not available in the Eurostat databases. Including all of the WB countries would provide a better insight into the differences concerning the ICT development level among these countries, and also into enterprises’ practice of the Internet and the websites usage.

Acknowledgments

This work has been fully supported by the Croatian Science Foundation within the project STRENGTHS (project no. 9402).

References


