INFORMATYKA EKONOMICZNA BUSINESS INFORMATICS 1(39) • 2016

ISSN 1507-3858
e-ISSN 2450-0003

Dorota Celińska
Uniwersytet Warszawski
e-mail: dcelinska@wne.uw.edu.pl

WHY DO USERS CHOOSE OPEN SOURCE SOFTWARE? ANALYSIS OF THE NETWORK EFFECT

DLACZEGO UŻYTKOWNICY WYBIERAJĄ OPROGRAMOWANIE *OPEN SOURCE*? ANALIZA EFEKTU SIECIOWEGO

DOI: 10.15611/ie.2016.1.01
JEL Classification: L17, L86, C38, D12

Summary: This article analyses the phenomenon of using the Open Source software. The aim is to investigate and verify the existence of a positive direct network effect that characterizes usage of the Open Source software. The multivariate probit model is applied to extract factors motivating end users to the usage of the Open Source software. Special attention is paid to the demographic characteristics of the end users, as well as to the impact of the end users' peers, such as family, work and school, on using the Open Source software. The results confirm the research assumptions. It is found that the actual choice of various Open Source software groups is correlated and thus cannot be analyzed separately. This supports the hypothesis about the existence of the network effect derived from the complementary services and dependencies among applications. The obtained results are representative for the Polish population aged 13-54.

Keywords: Open Source, software, end user, network effect, multivariate probit, lock-in effect.

Streszczenie: W artykule przeanalizowano zjawisko korzystania z oprogramowania *Open Source*. Celem było zbadanie i weryfikacja istnienia pozytywnego, bezpośredniego efektu sieciowego charakteryzującego oprogramowanie *Open Source*. Do określenia czynników motywujących użytkowników końcowych do korzystania z oprogramowania *Open Source* wykorzystano model *multivariate probit*. Szczególną uwagę poświęcono zarówno demograficznym charakterystykom użytkowników końcowych, jak i wpływowi otoczenia: rodziny, współpracowników czy szkoły. Wybór poszczególnych grup oprogramowania *Open Source* jest skorelowany, przez to nie można go analizować pojedynczo, co nie daje podstaw do odrzucenia hipotezy o istnieniu efektu sieciowego pochodzącego z usług komplementarnych i zależności między programami. Uzyskane wyniki są reprezentatywne dla populacji Polski w wieku 13-54 lat.

Słowa kluczowe: *Open Source*, oprogramowanie, użytkownik końcowy, efekt sieciowy, *multivariate probit*, efekt *lock-in*.

1. Introduction

The Open Source software constitutes a software class with its source code made available publicly. The license allows the end users to study, modify, and distribute the source code to anyone and for any purpose. In recent years there has been a rapid growth in the number of organizations that choose the Open Source licenses for their products. Even enterprises associated with typically commercial activities such as Samsung, Google or Microsoft tend to release large amounts of code for the needs of Open Source projects. This makes the Open Source one of the most discussed topics in analysis of software and information technology markets [Gallego, Luna, Bueno 2008; Celińska 2014].

The economic literature considering the Open Source phenomenon is limited to the supply side of the software market only [Crowston, Wei, Howison 2012]. The existing empirical research focuses on extracting the determinants of the Open Source license choice made on an enterprise (or a project) level [Bonaccorsi, Rossi 2003; Lerner, Tirole 2005; Koski 2007]. Additionally, a great amount of research has been conducted on explaining the incentives that motivate the developers to contribute to the Open Source projects [Hars 2002; Ye, Kishida 2003; Krishnamurthy, Ou, Tripathi 2014].

The aim of this article is to verify the existence of a positive direct network effect that characterizes usage of the Open Source software. To achieve this goal, a multivariate probit model is applied to the data obtained from a web-survey of Polish Internet users. The term "end users" signifies individual persons using legal software copies. Our main findings prove the existence of positive, and statistically significant, direct network effects while choosing the Open Source software. We also notice the significant dependencies among choices of groups of the Open Source software made by the end users.

The article is organized as follows. In Section 2 we give a short review of the existing empirical background. The data source is presented in Section 3. In the fourth section we formulate the research hypotheses and introduce the empirical model. In Section 5 we present the basic statistical characteristics of the sample, and the results of the conducted analysis are presented in Section 6. Section 7 concludes the research.

2. Network effect and end users of the Open Source software in recent research

The network effect is an externality that arises when consumption of usage of goods or services depends on the number of consumers [Church, Gandal, Krause 2008]. The positive network effect occurs when a growing number of consumers use the particular product − then it becomes more and more valuable for them. Therefore the consumer's utility function increases with the number of other users that choose

this product. The externalities arising from the network effect turn out to be among the most important economic factors stimulating the end users to use a particular software.

Network effect externalities may be classified as *direct*, *derived from complementary services* and *indirect*. End users are mainly affected by the first kind of externality. The classic example is the exchange of files. A simple, and apparently inevitable, outcome of the network effect is the so-called lock-in effect. This means that as soon as the software is treated as the standard, a significant increase in the supply of complementary services (e.g. applications, maintenance) is observed. If the software product gains a significant market share, the consumers have even more incentives to use it [Bonaccorsi, Rossi 2003]. The indirect network effect arising from the usage of the Open Source software is reported by Popovici [2007]. Open Source developers tend to integrate among themselves within the community, which stimulates the production of source code. The measure that is often used to assess the quality of the source code is called the bug resolution rate. In the case of the Open Source code, bugs are solved faster when the number of developers contributing to the project increases. Therefor developers' behaviour results in the higher quality of the delivered product, which indirectly affects consumers.

Gallego et al. [Gallego, Luna, Bueno 2008], identified the variables and factors that have a direct effect on individual attitudes towards Open Source software adoption. They utilised a multiequation model of the perceived usability, perceived ease of use and usage behaviour of Open Source software. Social influence in the form of the impact of users' peers and environment was also included. Those variables can be seen as the proxy for the direct network effect. The perceived ease of use and perceived usability significantly and positively affect the intention of using the Open Source software. On the other hand, the impact of users' peers and environment was not statistically significant, therefore the positive network effect was not found.

The demographic structure of active users-developers of the Open Source software was scrutinized during the Maastricht University 2002 project [Flossproject 2002]. The results were derived from a survey research of 2784 Open Source developers. The conclusion was drawn that the community of active Open Source users mainly consists of young males (98% of the sample, average age of 27, median age of 26) of high expertise in IT (70% of those surveyed with higher education, 83% employed in the ICT sector, or studying a related subject). Similar conclusions regarding the demography of Open Software users were obtained through a survey research known as *The Hacker Survey*, conducted by the Boston Consulting Group in 2001 [The Hacker Survey 2001]. The sample was created based on the answers of Sourceforge.com users, a site that gathers Open Source projects. Two groups of Open Source programmers were identified. The first one originated from Open Source projects at an early stage of development (alpha and beta), which consisted of a random sample of 10% of all applications listed under those stages in Sourceforge.

com repository. From this sample, 526 completed questionnaires were gathered (as compared to 1648 Open Source developers surveyed, a response rate of 34%). The second group covered mature Open Source projects assigned with at least one developer, 169 completed questionnaires were obtained (as compared to 573 people surveyed, a response rate of 30%) from this group. Relying on these data, it was found that active Open Source users are usually males (98% of the sample, average age of 30). Additionally, the low number of women among Open Source software users was confirmed in Kuechler's [2012] research.

The number of Open Source licensed applications steadily increases and covers applications designed for various groups of end users. The basic classification of Open Source end users can be derived from consumers' needs and their computer skills. According to Sen [2006], end users of the Open Source software can be divided into six separate groups: *advanced users*, *developers*, *desktop users*, *quality engineers*, *system administrators* and *others*. Some of the software categories (such as development tools) are dominated by this type of licensing, while in others (e.g. games) Open Source seems to be a minority. This may cause possible differences in the significance and strength of the network effect within the group of end users.

Gagné [2004], divides the end users of software into three separate groups. The first one includes *advanced users* with unique needs. They are familiar with the operating system GNU/Linux and are not afraid of changing their distribution of operating system. The second group consists of *average users* that need mostly desktop and office tools like text processing editor, spreadsheet or web browser. They have no technical skills in programming. The third group embraces people who need the not-Open Source applications that can hardly be replaced by the Open Source substitutes in their *professional* work.

Another division of Open Source users can be found in [Gacek, Arief 2004]. There are *passive users* that do not commit any patches to the projects and *active users*. Additionally, *active users* may be divided into a group of *developers* and a group of *non-developers*. *Non-developers* report issues about bugs and suggest new functionalities, but do not contribute to the source code. Finally, *developers* mean both *codevelopers* who support the development of an application and *coredevelopers* who are decisive about the way project is maintained.

3. Data source

We utilized a unique data based on an online web-survey data[1]. The target population was Internet users from Poland aged 13-54. The age was limited to this range in order to avoid potential bias of the underrepresentation of elderly people among Internet users. Data was collected from August 2012 to February 2013. The survey was preceded by a pilot survey in July 2012.

[1] The questionnaire is available on request.

A total of 1713 completed questionnaires were received from among 2412 responses, which constituted 71% of the number of questionnaires received. There were two subsets of the respondents: those surveyed through email-sent links and those visiting the survey web page on their own. Hence, it is not possible to calculate the total response rate for the survey. For the emailing subset of surveyed people, the response rate was 45%. After dropping observations that were out of the age range, the final sample consisted of 1694 responses.

It should be emphasized that the sample was not designed to be representative. To generalize the results, we decided to apply frequency weights. As the result of this correction, the distributions of the variables describing *age*, *gender*, *size of the place of residence* and *educational level* of end users became similar to those characterizing Polish society.

4. Research hypotheses, model and design

We stipulate that the Open Source software is characterized by a direct network effect that has a positive impact on the probability of becoming an Open Source software user as is reported by Bonaccorsi and Rossi [2003], and Popovici [2007]. Moreover, we expect that the actual choice of various Open Source software groups is correlated. Our analysis differs from the previous ones in various dimensions. Firstly, the behavior of the end users and their declared choices are analyzed. Secondly, the analysis is not limited to the Open Source developers and finally, the network effect derived from the complementary services and dependencies among applications is taken into consideration.

The respondents were asked about usage of eight separate groups of Open Source software in the survey. Since it could be difficult for the respondents to recall all of the applications they use, there were predefined sets of applications belonging to particular groups of interest of this study. Those were:

- *Office* applications: Libre/Open Office, Mozilla Firefox, Thunderbird, Chromium (Google Chrome), Claws Mail, 7zip, Pidgin, Android.
- *Hobby* applications: GIMP, Inkscape, Audacity, Amarok, VLC, Audacious, Rhytmbox, Blender, Mplayer.
- *Systems*: GNU/Linux, BSD, Open Solaris.
- *Websites* tools: Roundcube, SquirellMail, Drupal, Wordpress, Joomla, Piwik/OWA, Mediawiki, PhpMyAdmin.
- *Server* applications: Apache, Nginx, lighttpd, Filezilla, PuTTY, MySQL, PostgreSQL, MongoDB, Virtualbox.
- *Games*: Open Arena, Enemy Territory, 0AD, freeciv, Teeworlds, Tuxracer/SuperTux, Battle for Wesnoth.
- *Development* tools: Eclipse, Netbeans, Geany, Emacs, gedit, ViM, GCC, CLISP/SBCL, GHC/Hugs, Git, bazaar, Mercurial, SVN, PHP, Perl, Django, Python, Ruby.
- *Science*: Maxima, R, Octave, arduino, gretl, gnuplot, Weka.

The two most prominent groups of software used by respondents were *office* applications and operating *systems* (both around 50 percent of the sample). Another four groups of software were declared by 30 percent of each respondents. They described usage of *websites* tools, *server* applications, *development* tools and *games*. The last two groups of software: *science* and *hobby* applications were indicated by no more than 10 percent of end users.

The declared usage rate for software groups *websites*, *systems*, *servers* and *development* was too high in relation to the usage rate of those groups declared by the population of Poland aged 13 to 54 [Central Statistical Office of Poland 2012]. According to the CSO research, in 2011 the percentage of people capable of creating computer programs in Poland was three times lower than in this research sample, and the share of people creating Internet websites twice lower [Central Statistical Office of Poland 2012]. Unfortunately, the quoted report lacks the data that could be used as a reference to the usage rate of *science*, *games* and *servers* software groups. However, the distributions of variables in the sample are consistent with intuition. Programs from the *science* group are characterized by extraordinary specialized econometric-statistic application, which explains the low quantity of end users in the sample and population. The declared usage rate of *servers* is similar to the declared usage rate of website tools.

The diverse nature of applications adopted by the Open Source software end users prompted us to apply the 8-equation multivariate probit model. Probit models form the family where the dependent variable is categorical. The binary probit model is used to estimate the probability of binary response based on one or more independent variables. In our case, the dependent variables have a dummy nature and denote usage of the indicated Open Source group. The groups of applications were chosen from the most popular categories of Open Source software applications (according to the Sourceforge.net) and the categories of the Open Source applications mentioned by Sen [2006]. The multivariate probit is a kind of econometric model which is usually adequate when there is a suspicion of the problem of correlated errors among the equations of the model [Chib, Greenberg 1998]. In our analysis the following problem occurred: it is likely that the end user of Open Source software from one particular group would indicate also other groups and the choices of particular groups are not independent. In this case the multivariate probit model should be used instead of estimating separate binary probit models.

The first equation describes the probability of being an end user of *office* applications. In the second equation we estimate the probability of being an end user of *hobby* applications. The third equation describes the end users of Open Source operating *systems*. The probability of choosing Open Source tools designed for *websites* is calculated in the fourth equation. The fifth equation deals with the problem of choosing the *server* tools. In the sixth equation we estimate the probability of choosing Open Source *games*. End users of *development* tools are described by

the seventh equation. Finally, the eighth equation is devoted to the declared usage of *science* applications.

To uncover which factors determine the probability of using a certain group of the Open Source software, we chose two sets of explanatory variables. They consist of both sociodemographic and computer related characteristics. The first one encompasses *age* of a respondent and its square, *gender* dummy with 1 denoting man, *town* size, level of completed *education* and a dummy variable denoting being a *student*. The second set consists of computer related variables and is designed to capture the possible network effect.

To examine the existence of direct network effect for each of the software groups three variables were constructed. The variable *others* described the declared state of the respondent's knowledge about usage of the particular Open Source software in their environment. This variable was based on the responses to the question: "Do you associate other people with using the applications mentioned below". Within the eight aforementioned software groups the responses were valued: "no" – –1; "I do not know" – 0; "yes" – 1. Then, for every respondent the average of their valued responses within the software groups was calculated. That allowed for the creation of eight independent group-specific variables. The other view on direct network effect was captured by variables *source-friends* and *source-work*. The variable *source-friends* approximated the declared level of friends and family recommendations' impact on using the applications from the Open Source software group, meaning which applications used by the respondents are the ones that were previously recommended by their families and friends. The values of this variable were based on the responses to the question: "What was the origin of your knowledge about the application". Within the aforementioned software groups, the answer "recommendation of family or friends" was valued 1, and others were valued 0. Subsequently, the average of the valued responses of the respondent was calculated within the software groups. This denoted the fraction of the applications recommended by friends and family among all of the analyzed applications. Variable *source-work* covered the declared level of usage of the Open Source software groups at respondent's work, school or university, meaning which applications used by the respondents are the ones that were compulsory at educational units or work. This variable was constructed in the analogous way as *source-friends*, but the basic question was: "Where do you mostly use the application". The counted answer was "at work/school/university". Additionally, the squares of these three variables were included in the model.

5. Preliminary analysis

In Table 1 we present the basic descriptive statistics of variables *age* and *others* group. Left asymmetry is observed for the *others* group (average values are lower than median values). In general, respondents declare no knowledge about other people using a particular group of software. However, it can be noticed that respondents

very frequently declare knowledge about other people using *office* and *hobby* related software. The declaration regarding other software groups is considerably less frequent. Less than half of the respondents point at those groups. The least frequently indicated group is science applications.

Table 1. Basic descriptive characteristics of variables concerning the impact of age and direct network effect

Variable	Minimum	Maximum	Average	Median	Percent of values > 0
Age	13	54	33.50	33	—
Others_office	–1	1	0.29	0.38	77
Others_hobby	–1	1	–0.09	0.11	47
Others_systems	–1	1	–0.17	0	34
Others_websites	–1	1	–0.298	0	25
Others_servers	–1	1	–0.27	0	30
Others_games	–1	1	–0.38	0	14
Others_development	–1	1	–0.39	0	21
Others_science	–1	1	–0.46	0	7

Source: own analysis based on data from web-survey.

In Table 2 we present the basic descriptive statistics for source-work group. These variables are used as a proxy for compulsory usage of applications at work, school or the university. This kind of network externality may be connected with the lock-in effect.

Table 2. Basic descriptive characteristics of variables concerning usage of Open Source software at work/school/university

Variable	Minimum	Maximum	Average	Median	Percent of values > 0
Source-work_office	0	1	0.14	0	47
Source-work_hobby	0	1	0.05	0	23
Source-work_systems	0	1	0.09	0	22
Source-work_websites	0	1	0.05	0	19
Source-work_servers	0	1	0.08	0	24
Source-work_development	0	1	0.04	0	22
Source-work_science	0	1	0.02	0	5

Source: own analysis based on data from web-survey.

According to the data presented in Table 2, it can be observed that the distribution of independent variables from groups *source-work* is characterized by a slight right

asymmetry (average values are higher than median values). This means that the sample is characterized by a group of users that were encouraged to use specific Open Source software at work. The median value equal to the minimum value may indicate that at least half of the users in the sample were motivated to use a particular group of Open Source software by other factors than the compulsory usage. In general, the respondents declared the usage of software from all groups. The two most frequently obligatorily used groups of software was *office* applications and operating *systems*. This result was consistent with the results reported by CSO [Central Statistical Office of Poland 2011] that the most frequently used groups of Open Source applications in Polish enterprises were office applications together with web browsers and operating systems.

In Table 3 we present the basic descriptive characteristics of *source-friends* group. These variables were indicating what fraction of applications being used by the respondents are the ones that were recommended by their relatives and friends.

Table 3. Basic descriptive characteristics of variables concerning impact of the family and peers

Variable	Minimum	Maximum	Average	Median	Percent of values > 0
Source-friends_office	0	0.75	0.09	0	48
Source-friends_hobby	0	0.56	0.04	0	28
Source-friends_systems	0	1	0.06	0	17
Source-friends_websites	0	0.5	0.01	0	5
Source-friends_servers	0	0.67	0.02	0	10
Source-friends_games	0	0.57	0.01	0	7
Source-friends_ development	0	0.5	0.01	0	7
Source-friends_science	0	0.43	0.002	0	1

Source: own analysis based on data from web-survey.

We can state that the distributions of *source-friends* group, except from *source-friends_systems* and *source-friends_office*, are characterized by a minor right asymmetry. This means that the sample contained users that were slightly more often encouraged to use the Open Source software by their family and friends than other factors. Strong diagonal right asymmetry can be seen in the case of independent variables of *source-friends_systems* and *source-friends_office*. The value of median equal to the minimum value may suggest that at least half of the users in the sample were motivated rather by other factors than the advice of their relatives and peers.

In cases of discrete variable *town*, users from cities of over 200 thousand inhabitants constituted 20.5% of the sample, 7.8% of users were from cities of 100 000-200 000 inhabitants, 19.5% of users were from cities of 20 000-100 000 inhabitants, 12.9% of users were from cities below 20 000 inhabitants. People from the rural regions and

villages constituted 39.3% of the sample. The obtained distribution of the sample, in respect to the size of the place of residence, is consistent with the distribution of the size of place of residence taken from the report of the CSO Statistical Yearbook for 2012.

As much as 51.4% of the surveyed users had elementary educational level or lower. Secondary educational level was declared by 31.6%. At least first degree higher educational level was represented by 17% of the sample. The obtained distribution in the sample, in respect to the level of education, is consistent with the distribution of the level of education taken from the report of the CSO Statistical Yearbook for 2012.

After applying the frequency weights, students constituted 13.94% of the sample, and women 49%. The obtained distribution in the sample, in respect to the gender, is consistent with the distribution of the gender taken from the report of the CSO Statistical Yearbook for 2012.

6. Results of analysis

Firstly, we analyze the coefficients of correlations among equations in the model, called *rhos*. All of the *rhos* coefficients estimating the strength of dependence of the pairs of equations were found to be significantly different from 0 (p-value: $0.00 < 0.05$), apart from the correlation coefficient between the *website* equation and *hobby* equation (p-value: $0.52 > 0.05$). The LR test of joint insignificance rejected the null hypothesis (p-value: $0.00 < 0.05$). This means that the users' decisions about choosing the groups of the Open Source software are not independent, which confirms the validity of the model specification. The values of the obtained correlation coefficients are shown in Table 4.

Table 4. Estimated correlations of the equations in the multivariate probit model

X	Office	Hobby	Systems	Websites	Servers	Games	Development	Science
Office	1	0.19	0.39	0.20	0.16	0.15	0.30	0.38
Hobby	0.19	1	0.46	0.02	0.31	0.20	0.38	0.38
Systems	0.39	0.46	1	0.39	0.53	0.25	0.60	0.43
Websites	0.20	0.02	0.39	1	0.40	0.26	0.36	0.19
Servers	0.16	0.31	0.53	0.40	1	0.22	0.60	0.16
Games	0.15	0.20	0.25	0.26	0.22	1	0.37	0.19
Development	0.30	0.38	0.60	0.36	0.60	0.37	1	0.28
Science	0.38	0.38	0.43	0.19	0.16	0.19	0.28	1

Source: own analysis based on data from web-survey.

The coefficients presented in Table 4 result from a tetrachoric correlation. There is a positive correlation aamong the end users' decisions about choosing all

of the analyzed groups of the Open Source software. This supports our research hypothesis. The strongest correlation occurs between the decision about using Open Source operating systems and the decision about using Open Source development tools (*rho* = 0.6). High correlation coefficient (*rho* = 0.6) characterizes also the relation between the decision about using Open Source development tools and the decision about using Open Source servers. Using Open Source operating systems highly correlates with all of the other groups of software. The weakest correlations occur in the case of using Open Source games.

The results of the conducted analysis in the form of odds ratios for multivariate probit model are presented in Table 5. Since the set of variables in every equation is similar, for a concise presentation every column contains the values coming from the equation for the dependent variable specified in the first row.

The insignificant variables were excluded from the analysis as they not differentiate among equations. The validity of this step was confirmed by the joint significance test of those variables[2].

To validate the functional form of the model, we performed the linktest. The result turned out to be statistically insignificant (p-values > 0.05), which means that the functional form of our model is correct.

The coefficients for the independent variables *others* and *others2* were always statistically significant, and quantitatively the odds ratios were greater than one. This means that the Open Source software is characterized by a direct network effect which supports our research hypothesis. If the end user knows that their peers and family use certain groups of the Open Source software, the probability of making a similar decision of using certain groups of the Open Source software increases. It can be also observed that this probability increases nonlinearly (statistically important variable *others2*). Similar results were obtained by Bonaccorsi and Rossi [2003].

The coefficients for *source-friends* and *source-friends2*, describing the influence of end users' peers or family on the probability of adopting the Open Source software, were always statistically significant with the exception of variable *source-friends* in the equation for *servers*. Family and friends' influence shows diminishing returns to scale in the case of equations for *science*, *systems*, *games* and *websites*. This means that recommendation given by end user's peers and family influences positively the probability of the Open Source software group usage only to a certain degree. The obtained results give no basis to reject the research hypothesis about direct network effect characterizing the using of Open Source software. This contradicts the results described by Gallego et al. [Gallego, Luna, Bueno 2008].

---

[2] The joint significance test for the variable *source-friends2* in the equation for *office*, student in the equation for *systems* and in the equation for *development*, *source-friends2* and *source-work* in the equation for *servers* and *others2* in the equation for *science* supported the insignificance hypothesis (*p*-value: 0.27 > 0.05). Those variables were excluded from the further analysis. All of the remaining independent variables turned out to be jointly significant (*p*-value 0.00 < 0.05).

Table 5. Estimated correlations of the equations in the multivariate probit model

	Office	Hobby	Systems	Websites	Servers	Games	Develop.	Science
Age	1.07***	0.92***	1.05***	1.04***	0.91***	1.05***	0.89***	1.17***
Age2	0.99***	1.01***	0.99***	0.99***	1.01***	0.99***	1.01***	0.99***
Town200-500	0.88*	1.04	1.95***	1.05	1.65***	1.21***	0.89	0.79***
Town100-200	1.12*	2.53***	1.35***	1.03	1.77***	0.86***	1.26***	0.84**
Town20-100	0.91*	2.77***	1.21***	1.67***	1.39***	0.68***	1.14**	0.84***
Town20-	0.77***	2.66***	0.90*	0.86***	0.81***	0.68***	0.70***	0.94
Town-village	0.79***	1.77***	1.26***	0.91***	0.84***	0.67***	0.54***	0.58***
Education-sec	1.77***	2.41***	1.30***	1.14***	2.25***	1.42***	2.16***	0.90**
Education-bac	1.72***	1.40***	1.23***	1.27***	2.23***	1.57***	2.92***	0.45***
Education-moa	1.93***	0.61***	1.62***	1.07***	3.32***	1.93***	2.16***	0.93
Education-post	1.19*	1.54***	2.10***	1.09	0.71***	1.73***	0.99	0.29***
Student	0.61***	0.35***	X	1.08**	0.66***	0.88***	X	2.29***
Gender	0.30***	0.59***	0.19***	0.73***	0.23***	0.34***	0.24***	0.42***
Others	4.26***	3.56***	4.01***	7.61***	2.27***	3.39***	4.44***	3.10***
Others2	4.22***	3.49***	1.58***	5.21***	1.36***	1.68***	2.44***	X
Source-friends	61.56***	0.01***	59.0***	1074.92***	24.29***	28.57***	32.35***	95.01***
Source-friends2	X	10427***	0.04***	0.001***	X	0.001***	–	0.001***
Source-work	0.04***	0.13***	51.0***	3.29**	X	–	20.95***	47.96***
Source-work2	16.58***	651.97***	–	69.54***	23.31***	–	–	–

* − variable significant at significance level 0.1, ** − variable significant at significance level 0.05, *** − variable significant at significance level 0.01, X − excluded variables,

Source: own analysis based on data from web-survey.

The impact of the variable *source-work*, describing a degree of the Open Source software usage by the end user at work or school, is ambiguous. In cases of equations for *office*, *hobby*, *websites*, and *servers* the influence of that variable on the probability of making the decision of the Open Source software usage is nonlinear. This means that the higher number of the Open Source software belonging to a certain group of the Open Source software used at the end user's work or school, the higher the probability of the user making the decision of using that group of the Open Source software. The resulting conclusions give no basis to reject the research hypothesis of the existence of a positive direct network effect associated with the Open Source software usage. This kind of externality can be also seen as the lock-in effect.

7. Conclusions

This article analyzed the determinants of using the Open Source software. For that purpose we applied a multivariate probit model to unique dataset from the web-survey. The main goal of the presented analysis was to investigate the existence of the direct network effect associated with the Open Source software usage. It was found that the actual choice of various Open Source software groups is correlated and cannot be analyzed separately. This finding supports the hypothesis about the existence of the network effect derived from the complementary services and dependencies among applications.

Our analysis differs from the previous ones in various dimensions. Firstly, the behaviour of the end users and their declared choices are analyzed. Secondly, the analysis is not limited to the Open Source developers. Also, the network effects derived from the complementary services and dependencies among applications are taken into account.

Apart from the main findings, this study also showed that the gender of the user is an important factor in determining the probability of becoming an Open Source user. The analysis confirmed stronger than linear influence of user's age. There is a positive impact of achieving higher educational levels and living in bigger cities on using the Open Source software by the end user. Being a student has no statistically important influence in the case of choosing systems and development tools. The obtained results are consistent with common sense and intuition. Due to the pioneering and exploratory nature of this research, we find the obtained results satisfactory.

The obtained results are representative for the Polish population aged 13-54. They imply that the peers have a positive influence on choosing the Open Source software, which stands in contradiction with the results obtained by Gallego et al. [Gallego, Luna, Bueno 2008] that friends have no impact on the perceived usability of an application that drives the user's decision towards using that application. On the other hand, our findings were consistent with those described by Bonaccorsi and Rossi [2003], and Popovici [2007], which stated that an increasing number of people using the Open Source software motivates end users to use Open Source licensed applications.

References

Bonaccorsi A., Rossi C., 2003, *Why Open Source software can succeed*, Research Policy, 32, pp. 1243--1258.

Bonaccorsi A., Rossi C., 2006, *Motivations to take part in the Open Source movement*, Knowledge, Technology & Policy, vol. 18, no. 4, pp. 40-64.

Celińska D., 2014, *Użycie oprogramowania Open Source – co poza gift economy?,* Ekonomia, 37, pp. 5-25.

Central Statistical Office of Poland, 2012, *Społeczeństwo informacyjne w Polsce. Wyniki badań statystycznych z lat 2007-2011*, http://stat.gov.pl (18.03.2013).
Chib S., Greenberg E., 1998, *Analysis of multivariate probit models*, Biometrika, 85, pp. 347-361.
Church J., Gandal N., Krause D., 2008, *Indirect network effects and adoption externalities*, Review of Network Economics, vol. 7, no. 3, pp. 1-22.
Crowston K., Wei K., Howison J., 2012, *Free/libre Open Source software development: What we know and what we do not know*, ACM Computing Surveys, vol. 44, no. 2.
Flossproject, 2002, http://flossproject.org (4.12.2012).
Gacek C., Arief B., 2004, *The Many Meanings of Open Source*, IEEE Software, IEEE Computer Society, vol. 21, no.1, pp. 34-40.
Gagne M., 2004, *Moving to the Linux Business Desktop*, Addison-Wesley Professional.
Gallego M.D., Luna P., Bueno S., 2008, *User acceptance model of Open Source software*, Computers in Human Behaviour vol. 24, pp. 219-9–2216.
Gwembu K.L., Wang J., 2010, *Seeing eye to eye? An explanatory study of Free Open Source software users' perceptions*, The Journal of Systems and Software vol. 83, pp. 2287-2296.
Hars A., 2002, *Working for free? – mMotivations of participating in Open Source projects*, International Journal of Electronic Commerce, vol. 6 (3, Spring).
Koski H., 2007, *Private-collective software business models: coordination and commercialization via licensing*, Review of Economic Research on Copyright Issues, vol. 4, pp. 47-61.
Krishnamurthy S., Ou S., Tripathi A. K., 2014, *Acceptance of monetary rewards in open source software development*, Research Policy, vol. 43, pp. 632-644.
Kuechler V., 2012, *Gender Differences in Early Free and Open Source Software Joining Process*, Proceedings on Eighth International Conference on Open Source Systems.
Lerner J., Tirole J., 2005, *The Scope of Open Source Licensing*, Journal of Law, Economics and Organization, vol. 21, pp. 20-56.
Popovici I., 2007, *The determinants of Open Source quality: An empirical investigation*, Economics Research Working Paaper Series. Paper 49, pp. 1-39.
Sen R., 2006, *Open Source Software Development Projects: Determinants of Project Popularity*, EERI Research Paper Series, no. 2.
The Hacker Survey, 2001, http://bcg.com/opensource (17.12.2012).
Ye Y., Kishida K., 2003, *Toward an understanding of the Motivation of Open Source Software Developers*, Proceedings of International Conference on Software Engineering.