University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 5/2015 (153)

Dorota Celińska
Mirosława Lasek

Why do users choose Open Source software? Analysis of the network effect

Warsaw 2015

University of Warsaw
Faculty of Economic Sciences

Why do users choose Open Source software? Analysis of the network effect

Dorota Celińska
Faculty of Economic Sciences
University of Warsaw
e-mail: dcelinska @wne.uw.edu.pl

Mirosława Lasek
(post mortem)
Faculty of Economic Sciences
University of Warsaw

Abstract

This article analyses the phenomenon of using the Open Source software. Its aim is to verify the existence of a positive direct network effect that characterizes using of the Open Source software. The multivariate probit model is used to extract factors motivating users to the usage of the Open Source software. Special attention is paid to demographic characteristics of users, as well as to the impact of users' acquaintances, such as family, work and school on using the Open Source software. The results of the conducted analysis confirm our research.

Keywords:
Open Source, software, source code, end user characteristics, network effect, multivariate probit, motivation

JEL:
L17, L86, C38, D12

Introduction

The Open Source software constitutes a software class with publicly accessible source code that can be modified by users. In recent years a rapid growth in the number of organizations that choose the Open Source licenses for their products is noticed. 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 et al., 2008, Celińska, 2014].

The economic literature considering the Open Source phenomenon is limited to a supply side only [Crowston et al., 2012]. The existing empirical research focuses on extracting the determinants of the Open Source license choice made on a firm (or a project) level [Bonaccorsi and Rossi, 2003; Lerner and 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 and Kishida, 2003; Krishnamurthy et al., 2014].

Despite the large number of articles analyzing the Open Source software (e. g. Bonaccorsi and Rossi [2006], Gwembu and Wang [2010]), the reasons that push end users towards using Open Source software have hardly been a target of quantitative economic analysis so far. Through this article, we will try to fill the gap and take a look into a demand side.

The aim of this article is to verify the existence of a positive direct network effect that characterizes using of the Open Source software. To achieve this goal, multivariate probit model is applied to the data obtained from a web-survey of Polish Internet users. The term "users" signifies the "end users", i.e. 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 statistically significant dependence among choices of groups of the Open Source software made by the end users.

The article is organized as follows. In the section 1 we give a short review of the existing empirical background. The data source is presented in the section 2. In the third section we formulate the research hypotheses and introduce the empirical model. In section 4 we present basic statistical characteristics of the sample, and the results of the conducted analysis are presented in the fifth section. The Section 6 concludes the research.

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

The network effect occurs when more and more 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. Computer applications 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 a standard, a significant increase in the supply of complementary services (e.g. applications, maintenance) is observed. If a software gains a significant market share, the consumers have even more incentives to use it [Bonaccorsi and 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 a source code. The measure, that is often used to assess the quality of the source code, is called the bug resolution rate. In case of the Open Source code, bugs are solved faster when the number of developers contributing to the project increases. Therefore developers' behaviour results in the higher quality of a delivered product, which indirectly affects consumers.

On the other hand, the results obtained by Gallego et al. [2008] did not report existing positive network effect while using multi-equation model. According to this research, the impact of users' acquaintances and environment, which can be a substitute for direct network effect, was not statistically significant when it comes to the perceived usability and ease of use of the Open Source software. The perceived ease of use and perceived usability significantly and positively affect the intention of using the Open Source software.

Demographic structure of active users-programmers 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 programmers. 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, with regard to 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 basing 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 a random sample of 10% of all applications listed under those stages in Sourceforge.com repository. From this sample 526 completed questionaires were gathered (as compared to 1648 Open Source programmers surveyed, response rate of 34%). The second group covered mature Open Source projects assigned with at least one programmer. From this sample 169 completed questionaires were obtained (as compared to 573 people surveyed, response rate of 30%). Relying on these data, it was found that active Open Source users are usually males (98% of the sample, average age of 30). Additionally, 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 users. The basic classification of Open Source users can be

derived from consumers' needs and their computer skills. According to Sen [2006], users of the Open Source software can be divided into 6 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 other (e.g. games) Open Source seems to be a minority. This may cause possible differences in significance and strength of the network effect within the group of users.

Gagné [2004] divides the users of software into three separate groups. The first one includes advanced users with unique needs. They are familiar with operating system GNU/Linux and are not afraid of changing their operating system's distribution. The second group consists of average users that need desktop and office tools, e.g. 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 partition of Open Source users can be found in [Gacek, 2004]. There are *passive users* that do not commit any patches to the project 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 the 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.

2. Data sources

We utilized a unique data based on an online web-survey data[1]. The target population was the 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 the Internet users. Data was collected from August 2012 to February 2013. The survey was preceded by a pilot survey in July 2012.

A total of 1713 complete questionnaires were received from all 2412 responses, which constituted the 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 45.1%. 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 use frequency weights. This resulted in the distributions of the variables describing *age*, *gender, size of the place of residence* and *educational level* of users, which was similar to those characterizing Polish society.

[1] The questionnaire is available on request.

3. 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 it 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 differ from the previous ones in various dimensions. Firstly, the behaviour of the end users and their declared choices are analysed. Secondly, the analysis is not limited to the Open Source developers. And finally, the network effect derived from the complementary services and dependencies between applications is taken into consideration.

The respondent were asked about usage of 8 separate groups of Open Source software in the survey. The two most prominent groups of software used by respondents were *office* applications and operating *systems* (both around 50 percent of the sample). The *office* applications included: Libre/Open Office, Mozilla Firefox, Thunderbird, Chromium (Google Chrome), Claws Mail, 7zip, Pidgin, Android; while *systems* covered usage of GNU/Linux, BSD and Open Solaris. Another four groups of software have been declared by 30 percent of respondent each. They described usage of *websites* tools, *server* applications, *development* tools and *games*. *Websites* tools encompassed: Roundcube, SquirellMail, Drupal, Wordpress, Joomla, Piwik/OWA, Mediawiki, PhpMyAdmin. The *server* applications belonged to the following group: Apache, Nginx, lighttpd, Filezilla, PuTTY, MySQL, PostgreSQL, MongoDB, Virtualbox. *Development* tools embraced Eclipse, Netbeans, Geany, Emacs, gedit, ViM, GCC, CLISP/SBCL, GHC/Hugs, Git, bazaar, Mercurial, SVN, PHP, Perl, Django, Python and Ruby. The group g*ames* included: Open Arena, Enemy Territory, 0AD, freeciv, Teeworlds, Tuxracer/SuperTux and Battle for Wesnoth. The last two groups of software: *science* and *hobby* applications were indicated by no more than 10 percent of users. The group *science* covered the usage of Maxima, R, Octave, arduino, gretl, gnuplot and Weka and the *hobby* applications belonged to the group: GIMP, Inkscape, Audacity, Amarok, VLC, Audacious, Rhytmbox, Blender, Mplayer.

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 [CSO, 2012]. According to the CSO research, in 2011 the percentage of people capable of creating computer programs in Poland was 3 times lower than in this research sample, and the share of people creating Internet websites 2 times lower [CSO, 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 the studied sample are consistent with intuition. Programs from the *science* group are characterized with extraordinary specialized econometric-statistic application, which explains low quantity of users in population. The declared usage rate of *servers* is similar to the declared usage rate of *websites*.

The diverse nature of applications mentioned by users of the Open Source software prompted us to apply 8-equations multivariate probit model. This kind of econometric model is usually used when there is a suspicion of the problem of correlated errors between the equations of the model

[Chib, Greenberg, 1998], which might occur in our analysis. It is likely that the user of Open Source software from one particular group would indicate also another groups. In each equation the dependent variable has a dummy nature and denotes usage of the indicated Open Source groups. The groups 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 first equation describes the probability of being a user of *office* applications. In the second equation we estimate the probability of being the user of *hobby* applications. The third equation describes the 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*. Users of *development* tools are described by the seventh equation. Finally, the eight equation is devoted to declared usage of *science* applications.

To uncover which factors determine the probability of used groups of the Open Source software, we chose a set of explanatory variables. This set consists of both sociodemographic and computer related characteristics. The first set encompasses *age* of a respondent and its square, *gender* dummy with 1 denoting man, *town* size, level of completed *education* and dummy variable denoting being a *student*. The second set of computer related variables was 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 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* aproximated 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 other 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". Additionaly, the squares of these three variables were included in the model.

4. Preliminary analysis

In Table 1, we present basic descriptive statistics of independent continuous 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 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 basic descriptive statistics for group of continuous independent variables *source-work*. These variables are used as a proxy for compulsory usage of applications at work, school or the university.

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

Variable	Minimum	Maximum	Average	Median
source-work_office	0	1	0.14	0
source-work_hobby	0	1	0.05	0
source-work_systems	0	1	0.09	0
source-work_websites	0	1	0.05	0
source-work_servers	0	1	0.08	0
source-work_development	0	1	0.04	0
source-work_science	0	1	0.02	0

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 larger group of users that were encouraged to use specific Open Source software at work. In general, the respondents declared the usage of software from all groups. The two most frequently obligatory used group of software was *office* applications and operating *systems*. This result was consistent with the results reported by CSO [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 basic descriptive characteristics of continuous independent variables *source-friends*. 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 the impact of the family and acquaintances.

Variable	**Minimum**	**Maximum**	**Average**	**Median**
source-friends_office	0	0.75	0.09	0
source-friends_hobby	0	0.56	0.04	0
source-friends_systems	0	1	0.06	0
source-friends_websites	0	0.5	0.01	0
source-friends_servers	0	0.67	0.02	0
source-friends_games	0	0.57	0.01	0
source-friends_development	0	0.5	0.01	0
source-friends_science	0	0.43	0.002	0

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

We can state that the distributions of independent variables of *source-friends*, except independent variables of *source-friends_systems* and *source-friends_office*, are characterized by a minor right asymmetry. This means that the sample consisted of 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 case of independent variables of *source-friends_systems* and *source-friends_office*.

In case of discrete variable *town*, users from cities of over 200 thousand of 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. 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 coming from the report of CSO Statistical Yearbook of the year 2012.

As much as 51.4% of the surveyed users exhibited the elementary educational level or lower. 31.6% of users declared the secondary educational level. 17% of the sample represented at least the first degree higher educational level. Obtained distribution of the sample, in respect to the level of education, is consistent with the distribution of the level of education coming from the report of CSO Statistical Yearbook of the year 2012.

In case of binary independent variables, 32.29% of surveyed users were students. 33.35% of sample consisted of women. After applying the frequency weights, students constituted 13.94% of the sample, and women 49%. Obtained distribution of sample, in respect to the gender, is consistent with the distribution of the gender coming from the report of CSO Statistical Yearbook of the year 2012.

5. Results of analysis

At first, we analyze the coefficients of correlations between equations in the model, called *rhos* coefficients. All of the *rhos* coefficients estimating the strenght 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 which confirms the validity of the model specification. The obtained correlation coefficients are shown in the 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 between the 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 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 the Table 5.

Table 5. The results of multivariate probit model (odds ratios).

	office	hobby	systems	websites	servers	games	development	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***
town 100-200	1.12*	2.53***	1.35***	1.03	1.77***	0.86***	1.26***	0.84**
town 20-100	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	2.92***
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	104279***	0.04***	0.001***	X	0.001***	--	0.001 ***

	office	hobby	systems	websites	servers	games	development	science
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***	–	--	–

Source: Own analysis based on the web-survey data.
* denotes variable significant at significance level 0.1
** denotes variable significant at significance level 0.05
*** denotes variables significant at significance level 0.01
X denotes excluded variables

The insignificant variables were excluded from the analysis as they not differentiate between equations. The validity of this step was confirmed by the joint insignificance test of those variables.

The joint significance test for the variable *source-friends2* in equation for *office*, *student* in the equation for *systems* and in the equation for *development*, *source-friends2* and *source-work* in equation for *servers* and *others2* in 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).

To check if the function form of the model is correct, 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 user knows that his acquaintances 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 independent variables *source-friends* and *source-friends2*, describing the influence of recommendation of an Open Source software group by users acquaintances or family on the probability of making the decision of the Open Source software usage, 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 case of equations for *science, systems, games* and websites. This means, that recommendation of the Open Source software by user's acquaintances and family influences positively the probability of making the decision of the Open Source software group usage only to a certain degree. 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. [2008].

The odds ratios for the variable *source-work*, describing a degree of the Open Source software usage by the user at work or school are always greater than zero. In case of equations for *office*, *hobby*, *websites*, and *servers* the influence of that variable on 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 is used at user's work or school, the higher the probability of user making the decision of using that group of the Open Source software. Resulting conclusions give no basis to reject the research hypothesis of existence of positive direct network effect associated with the Open Source software usage.

6. Conclusions

This article analyzed the determinants of using the Open source software. For that purpose we applied multivariate probit model to unique dataset from the web based 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 between applications.

Our analysis differ 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 between applications are taken into account.

Apart from main findings, this study also showed that 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 a 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 exploratory characteristic of this research, we find the obtained results satisfying.

The obtained results are representative for the Polish population aged 13-54. They imply that the acquaintances have positive influence on choosing the Open Source software, which stands in contradiction with results obtained by Gallego et al. [2008] that friends have no impact on 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 increasing number of people using the Open Source software motivates user to use Open Source licensed applications.

References

[1] Bonaccorsi A., Rossi C., (2003). Why Open Source software can succeed. *Research Policy* 32, 1243-1258.

[2] Bonaccorsi A., Rossi C., (2006). Motivations to take part in the Open Source movement. *Knowledge, Technology & Policy* vol. 18 n. 4, 40-64.

[3] Celińska D., (2014). Użycie oprogramowania Open Source – co poza gift economy? *Ekonomia* 37, 5-25.

[4] Chib S., Greenberg E., (1998). Analysis of multivariate probit models. *Biometrika* 85.

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 v(44,2).

[5] Gacek C., Arief B., (2004). The Many Meanings of Open Source. *IEEE Software, IEEE Computer Society*, vol(21).

[6] Gagne M., (2004). Moving to the Linux Business Desktop. Addison-Wesley Professional.

[7] Gallego M.D., Luna P., Bueno S., (2008). User acceptance model of Open Source software. *Computers in Human Behaviour* 24.

[8] 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* 83.

[9] Hars A. Working for free? – Motivations of participating in Open Source projects. *International Journal of Electronic Commerce*, 2002, v6(3, Spring).

[10] Koski H., (2007), Private-collective software business models: coordination and commercialization via licensing. *ETLA Discussion Papers, The Research Institute of Finnish Economy* 109.

[11] Krishnamurthy S., Ou S., Tripathi A. K., (2014). Acceptance of monetary rewards in open source software development. *Research Policy* 43, 632-644.

[12] Kuechler V., (2012). Gender Differences in Early Free and Open Source Software Joining Process. *Proceedings on Eighth International Conference on Open Source Systems.*

[13] Lerner J., Tirole J., (2005). The Scope of Open Source Licensing. *Journal of Law, Economics and Organization* 21, 20-56.

[14] Popovici I., (2007). The determinants of Open Source quality: an empirical investigation. *Florida International University Working Papers* 704.

[15] Sen R., (2006). Open Source Software Development Projects: Determinants of Project Popularity. *EERI Research Paper Series* 2.

[16] Ye Y., Kishida K., (2003). Toward an understanding of the Motivation of Open Source Software Developers. *Proceedings of International Conference on Software Engineering*.

[17] Główny Urząd Statystyczny, (2012). Społeczeństwo informacyjne w Polsce. Wyniki badań statystycznych z lat 2007-2011. [online 18.03.2013]
http://stat.gov.pl/cps/rde/xbcr/gus/nts_spolecz_inform_w_polsce_2007-2011.pdf

[18] *The Hacker Survey*, (2001). http://bcg.com/opensource [online 17.12.2012]

[19] *Flossproject*, (2002). http://flossproject.org [online 4.12.2012]

Faculty of Economic Sciences
University of Warsaw
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl