Paper to be presented at the

DRUID Society Conference 2014, CBS, Copenhagen, June 16-18

Organizing a Firm-Community Collaboration for growth: How to benefit from open source projects without hurting them

Dilan Aksoy-yurdagul
ESC Rennes School of Business
Department of Strategy and Innovation
daksoy@emp.uc3m.es

Francesco Rullani
LUISS Guido Carli
Management
frullani@luiss.it

Cristina Rossi-lamastra
Politecnico di Milano
Department of Management, Economics, and Industrial Engineer
rossi@mip.polimi.it

Abstract

This paper focuses on collaborations among firms and communities for obtaining better outcomes from open source projects. In specific, we aim to disentangle how organizational design may affect the performance of an OSS project. Project?s management model and employee involvement in the project are treated as potential mediators that may have an effect on the aforementioned relationship. The empirical analysis is undertaken on a sample of OSS projects hosted on the platform SourceForge.net from December 2006 to December 2008. The findings of this study are three fold. First, being directly involved in a project with a specific policy on OSS has a positive effect on project?s performance. Second, coordination by firm has a negative effect on performance. Third, admin as an employee on main duty does not have a direct positive effect on performance. However, it positively moderates the aforementioned relationship. The findings contribute to knowledge on benefits of collaborations between firms and OSS communities. Firm?s role in the complex coordination mechanism of the project on project?s success has been investigated. Managing the boundaries of collaborations is essential. Written rules and guidelines lead to fruitful joint development of software.

Jelcodes:O32,O31

Organizing a Firm-Community Collaboration for growth: How to benefit from open source projects without hurting them

ABSTRACT

This paper focuses on collaborations among firms and communities for obtaining better outcomes from open source projects. In specific, we aim to disentangle how organizational design may affect the performance of an OSS project. Project's management model and employee involvement in the project are treated as potential mediators that may have an effect on the aforementioned relationship. The empirical analysis is undertaken on a sample of OSS projects hosted on the platform SourceForge.net from December 2006 to December 2008. The findings of this study are three fold. First, being directly involved in a project with a specific policy on OSS has a positive effect on project's performance. Second, coordination by firm has a negative effect on performance. Third, admin as an employee on main duty does not have a direct positive effect on performance. However, it positively moderates the aforementioned relationship. The findings contribute to knowledge on benefits of collaborations between firms and OSS communities. Firm's role in the complex coordination mechanism of the project on project's success has been investigated. Managing the boundaries of collaborations is essential. Written rules and guidelines lead to fruitful joint development of software.

1. Introduction

Recently, innovation projects have "opened-up". External linkages are now increasingly used by firms to tap on additional resources to develop their innovation and knowledge creation processes. Research has shown that the typology of external sources a firm can benefit from may be very heterogeneous (Laursen & Salter, 2006). Universities (Laursen & Salter, 2004) users (von Hippel, 1988) and suppliers (Jayaram, 2008) are only part of the possible subjects involved by firms in their innovation projects.

A lively stream of research on this theme has recently shown that firms collaborate also with communities of users and developers (Dahlander, Frederiksen, & Rullani, 2008; Jeppesen & Frederiksen, 2006; Sproull, Dutton, & Kiesler, 2007), i.e. firms could innovate by engaging in community-based innovation projects (CIP). Collaborations with communities engender substantial benefits to firms by augmenting and complementing the scientific, artistic, or technological knowledge that they produce internally (Hargrave & Van de Ven, 2006).

The literature on new product development (Trott, 1998) has shown that specific features of the projects and of the innovation to be produced (Brun, Saetre, & Gjelsvik, 2009), as well as the environmental factors (Akgün, Byrne, Lynn, & Keskin, 2007), and the firm specific characteristics (Swink & Song, 2007) affect considerably the success of the project itself and its sustainability. Among these variables, a particularly relevant construct deals with the characteristics of the participants in the project. The relative heterogeneity (Giuri, Ploner, Rullani, & Torrisi, 2010) or familiarity (Akgün, Byrne, Keskin, Lynn, & Imamoglu, 2005) of the involved actors, their communication channels and the way in which interaction is organized (Schulze & Hoegl, 2006) are qualities and processes that clearly determine –to various extent- the project's capability to survive over time and the probability of a positive innovation outcome.

At present, communities producing Open Source software (OSS) within OSS projects figure prominently among users and developers communities. The OSS movement has recently experienced a radical metamorphosis: it has acquired commercially viable forms and has rapidly increased its economic importance (Dahlander, 2007; Fitzgerald, 2006). Software firms are getting increasingly involved in the OSS movement. Many firms, have adopted OSS-based business models (Dahlander & Magnusson, 2005; Gruber & Henkel, 2006) and have pursued to do business out of the OSS code that is freely downloadable from the Internet. Likewise,

software firms' participation in OSS projects[1] is currently a widespread phenomenon. Firms participate in the projects of collective software development created by OSS communities through varied participation strategies. Firms sponsor and even coordinate OSS projects (West & O'Mahony, 2008), pay their employees to develop and debug open code (Hars & Ou, 2002), reveal their own proprietary software code to OSS developers (Henkel, 2009), and so on.

There is important heterogeneity on how firms deal with OSS, since, when involved, some firms may have specific policies on OSS in order to establish the guidelines of collaboration. Projects, on the other hand, may vary largely by their governing models. While some are coordinated by a non-profit organization, some are coordinated by for-profit firms. In some cases there might be a formal leader in the project or the project as well may be coordinated informally. Several studies have explored the benefits from OSS projects' participation (e.g. Henkel, 2009) in order to disentangle which incentives drive for-profit firms to engage in this form of collective action (see Capra, Francalanci, Merlo, & Rossi-Lamastra, 2011). Another research stream has focused on the design of firms' collaborations within OSS projects. In this respect, it has been noted that mutual adaptations are needed. Specifically, contrary to what one can expect for innovation projects originating and developed in a more "usual" firm-controlled environment, for fruitfully participating in OSS projects firms must properly design their business models (e.g. Bonaccorsi, Giannangeli, & Rossi, 2006; Dahlander & Magnusson, 2005) and modify their internal organization (Colombo, Piva, & Rossi, 2010; Dahlander & Wallin, 2006) to be able to accommodate the specific peculiarities of a subject (the community) that is placed out of their control. At the same time, firms need to support the capabilities of OSS project communities to mobilize voluntary contributors (O'Mahony & Bechky, 2008; West & O'Mahony, 2008). In this line, a more recent research avenue investigates the impact of OSS projects' participation on firms' performance (Stam, 2009).

However, the lively debate on the theme, research on firms' involvement in community-based projects still leaves room for further investigations (Oreg & Nov, 2008; Roberts, Hann, &

[1] An OSS project is defined as "any group of people developing software and providing their results to the public under an Open Source license" (Evers, 2000). OSS projects are generally made available online in so called development platform, which are also repositories for the code itself. These are dedicated web sites that provide a suitable environment for software development and interaction between programmers. The largest and most well-known repository is SourceForge (http://sourceforge.net). At the beginning of January 2010, more than 380,000 OSS projects and 2 million users were registered at SourceForge.

Slaughter, 2006). Specifically, it is undoubtedly reasonable to expect that firm's role in project's coordination structure have diverse effects on projects' performance. This paper aims to fill this literature gap by answering the following research questions: 1) How does direct firm involvement with a specific policy affect the performance of a CIP? 2) How does coordination by firm affect the performance of a CIP? 3) Does the involvement of admin in the firm have a moderating role on the relationship between coordination by firm and CIP's performance?

The issue is of undoubted relevance. A crucial prerequisite for firms' external knowledge sourcing activities from OSS communities is the existence of a long-lasting knowledge flow from these peculiar external sources. The tension between the members of the community and the professional firm in governing the project may lead to inefficient coordination that eventually may dry up this flow and even challenge the very survival of the OSS project. Being inspired by profit-seeking motives, firms' role as the sole authority in governing the project run the risk of fatally undermine the core of OSS communities' activity, which is often based on values at odds with the firms' profit-oriented goals (O'Mahony & Bechky, 2008).

The findings of this study are three fold. First, being directly involved in a community with a specific policy has a positive effect on project's performance. Second, coordination by profit-oriented firm has a negative effect on performance. Third, admin as an employee on main duty does not have a direct positive effect on performance. However, it positively moderates the aforementioned relationship. The study contributes to our knowledge on benefits of collaborations between firms and communities. Firm's role in the complex coordination mechanism of a CIP and its effect on CIP's success is studied in an empirical setting. To our knowledge, there are no studies in the management literature that assesses the effect of firm policy, which describes how to work within the community, on the CIP's performance.

In the next section, we present the theoretical background for our empirical analysis. Then, we describe the data and methodology. After presenting the results, we offer a brief discussion of our findings and conclude.

2. Theoretical Background and Hypotheses

Firms are exerting significant efforts to engage in communities because they may receive several benefits in terms of project launch, development and realization (Henkel, 2006; von Hippel & von Krogh, 2003). Although literature has identified the potential benefits of engaging

in communities of users and developers (Dahlander & Wallin, 2006; West, 2003), heterogeneities in projects' success that might have been originated from firm involvement in these projects need to be investigated. In specific, we aim to link firm's role in the complex coordination mechanism of a CIP with the project's performance. Key drivers of sustaining a fruitful collaborative project are aimed to be investigated in depth.

While firms' efforts in allocating resources to a CIP may seem most likely to enhance the outcome of a project, it brings together many potential challenges. A critical error that a firm may commit when engaging in CIP is its failure in not clearly defining the role of those employees who are assigned to work on CIPs (Goldman & Gabriel, 2005). In most OSS projects, the members decide on what modules they are going to work on. Self-assignment of tasks by community members has its efficient side since each participant would effectively work in the task at maximum his or her productivity, exploiting a good fit between skills and tasks (Benkler, 2002). However, as the project evolves and involves several contributors (e.g. firm employees), coordination problem could arise since the allocation of contributors per task could be no homogenously distributed, with tasks overcrowded and tasks deserted. This "chaotic" style of development could undermine project effectiveness (Mockus, Fielding, & Herbsleb, 2000). In order to avoid such a problem in a joint project, a possible solution is to first have a detailed map of the skills and tasks available and then design the best work allocations among developers (Mockus et al., 2000). Clearly, assigning work activities among team members will increase a CIP's performance.

Undefined metrics to be used in measuring the success of the OSS project may as well lead to inefficiencies while the opposite has been aimed by the firm when getting involved in the project (Goldman & Gabriel, 2005). Failure in defining the metrics for the success can put a CIP under strain and will give rise to neglecting community activities and its well-being by employees of the firm (Pelled, Eisenhardt, & Xin, 1999). Employees of a firm should be well-informed about the potential benefits that the CIP will bring about to the firm so that they understand the importance of the OSS project and be concerned about the overall health of the community (Wagstrom, Mockus, Herbsleb, & Kraut, 2010). Defining the metrics used to measure the success of the project will ensure employee dedication to the project by simply stating that the project is not less important than any other development project that is done in-house (Goldman & Gabriel, 2005, p.266).

A firm could avoid these backfires if a firm's involvement in the project is explicitly guided by a formal policy, formally defined and made visible to the community members. A specific policy from one side defines ex-ante the role of the firm employees that join CIPs vis-a-vis the volunteer contributions from the community; it specifies how the CIP will contribute to firm's business goals without compromising CIP rules and legitimacy.

Hypothesis 1: Community-based innovation projects with a direct involvement of a profit-oriented firm will reach higher success when the firm participates with a specific policy.

However, some trade-off could always be present between profit-oriented aims and the motivations of an open community. Extracting financial benefits from CIPs contradicts, at the end, the same core values of OSS movement (O'Mahony, 2003). Since the balance between profit and non-profit oriented aims sometimes is not trivial, collaborations among parties can be difficult when the interests, goals, and practices of participants differ (O'Mahony & Bechky, 2008). When a firm joins a CIP, it is likely that it will market the project to attract as many volunteer developers as possible and assign its own full time employees to work on the project as well (Dahlander & Wallin, 2006; Henkel, 2009). Thus, the firm will most probably decide according to its strategic aims to assign paid employees to particular project tasks, especially to the ones that are critical for generating economic returns. This could generate conflicts among project's participants due to the free mindset of volunteers and regulated behavior of firm employees (Henkel, 2009; Lakhani & Wolf, 2005). In the long run, the heterogeneities originated from the diverse work disciplines, communication channels and skill sets of participants may have a negative effect on the overall productivity (Jackson et al., 1991; Star & Griesemer, 1989). The coordination figure of a CIP is indeed very important to set the coordination style. "A coordination style depends on a certain way how developers work on the source code, and more important, how they interact with each other" (Spaeth, 2005). A focal firm acting as the principal authority in a CIP may create a tension within the community, because volunteers know that the main coordinator behavior is dictated by profit maximization schemes, and could not be trusted in every critical decision. This could deter volunteer participation.

Even if the interests align, one thing is to say that for-profit organization could participate fruitfully in a successful CIP, another is to say that for-profit firms could lead and manage open

communities based on non-profit rules and motivations (Timmermans & Leiter, 2000). Firms' involvement could be also damaging due to the bureaucratic and restrictive nature of the firm that favor hierarchical division of labor (Daniel, Maruping, Cataldo, & Herbsleb, 2011). According to Rucht (2004), when firms lead CIPs it becomes hard to strike a balance between trust and control. While volunteers' participation is mainly driven by intrinsic motivations, firms may try to publicly manage CIPs like internal proprietary projects by neglecting community participation, by disabling democratic development of the project, and/or by adopting a clear free-riding approach (i.e. sneaking code) (Goldman & Gabriel, 2005). For-profit actions that are in contrast with the open access moral rule of CIPs will destroy the community's trust and discourage participation (Stewart & Gosain, 2006). Furthermore, disrupting the community will result in additional costs for the firm since some of the work load originally performed by volunteers may be abandoned (Wagstrom et al., 2010). Thus, we suggest that coordination by profit-oriented firm will have a negative effect on project's performance.

Hypothesis 2: Community-based innovation projects with a direct involvement of a profit-oriented firm will reach lower success when the firm participates as coordinator.

Drawing on social network theory it is argued that there are favorable network positions that enable firms to manage key resources without ownership in the legal sense (Burt, 1992; Wasserman, 1994). Dahlander and Wallin (2006) suggest that firms act strategically by claiming these favorable network positions in an OSS community to be able to gain access to privileged information. Key person in a professional network may act as a technological broker by combining the best of both worlds (Hargadon & Sutton, 1997) and build legitimacy, which can be achieved through proof of skillfulness and by providing help to other individuals in the community. The administrator of the project, who at the same time is an employee of the firm, might be a key person that supports and encourages volunteer participation by maintaining a close interaction with the community participants. However, it is important to what degree the admin of the project socially identifies himself/herself with the community and the organization he/she works for at the same time. Henkel (2009) suggests that developers' identification with the company they work for and with the community they participate has an effect on their development-specific attitudes and behaviors. Similarly, administrators' attitudes, when managing the project, may as well be affected by how they identify themselves with the firm and

with the community. Furthermore, due to the key position the administrator holds in the project, volunteer participation might decrease or increase depending on the administrator's behaviors throughout the project development process.

How an individual associates oneself with a social group (e.g. an organization, a profession) can be defined as his social identity (Tajfel & Turner, 1986). The degree to which an individual identifies oneself with the organization he/she works for and with the open source community have been considered as important moderators affecting the relationship between performance and coordination by profit-oriented firm (Daniel et al., 2011). If the admin of the project highly identifies himself/herself with the company (e.g. owner), the negative effect of coordination by firm on performance may worsen since his/her attitudes will favor company benefits primarily. On the other hand, if the admin of the project loosely identifies himself/herself with the community (e.g. employee, who has a main duty other than working on the project), the negative effect of coordination by firm on performance will not improve neither since the time spent as admin is not enough to generate a high identification with the community to seek community's interests (Gunz & Gunz, 2007).

We suggest that if the administrator of the project is an employee of the firm whose main duty is to work on the project, the negative effect of coordination by firm on the project's performance will decrease. The admin of the project, who at the same time is an employee of the firm, may offset the tension between the community and the firm by building legitimacy. Furthermore, a firm employee fully dedicated to work on the project as the admin will have a favorable position both in terms of financial resources and accessibility to the results and developments from in-house investments. The admin, who identifies himself/herself with the community as much as with the company he/she works for, may help the community develop more organically instead of pushing a strong corporate agenda that will seem suspicious to the community and thus, will positively moderate the relationship between coordination by firm and project's performance.

Hypothesis 3: The negative effect of the coordination role of a firm in a community-based innovation project will be less stringent when the administrator of the project has a double organizational identity.

3. RESEARCH METHODS

3.1. Sample and Data Collection

Our empirical analysis is undertaken on a sample of Java-based[2] OSS projects hosted on the platform SourceForge.net[3], which is the largest and most popular repository of community based OSS developments. Multiple sources of data have been used to gather information that enables us for grounded theory development and validation of theoretical constructs.

An online survey was run in late November 2007 to collect data on projects' coordination styles and firms' various involvement modes within these OSS projects (Capra et al., 2011). Specifically, the online survey was addressed to all the administrators of the 8,308 OSS projects written in Java and hosted on SourceForge.net as of November 2007 (see Capra et al., 2011 for further details). The survey includes questions that aim at identifying project characteristics, which are not possible to retrieve from the internet and also questions regarding the leading company, when there is firm involvement in the project. The questions related to firm involvement were more specifically about the role of the firm (e.g. sponsor, founder, directly involved with or without specific policy, simply acknowledges participation of its employees), the relation of the administrator of the project with the firm, and the management method of the project. We also gathered secondary data on each firm identified in this way using Internet searches and calling up directly the firms when needed.

The questionnaire was hosted on SurveyMonkey.com, a website specialized in online surveys. Each project's administrator was contacted via e-mail by sending him/her the URL of the questionnaire and a reference institutional home-page. In case of ambiguous or incoherent answers, she or he was re-contacted via e-mail and interviewed to clarify the goals of the questionnaire and obtain correct data. These interviews were also used to check whether the project information declared on SourceForge.net was correct. Before submitting the questionnaire to the selected administrators, a pilot test on a random sample of 195 administrators, stratified by project size, was run. This pilot test was carried out with several OSS developers and project administrators, who extensively discussed a preliminary version of the questionnaire. The final version accounts for the feedback and comments by the participants

[2] Java programming language is one of the most popular programming languages.
[3] cf. http://www.sourceforge.net

of this pilot stage. The survey was conducted starting from November, 27th 2007. After the first mailing, two follow-ups were performed, starting from December, 5th 2007 and December, 18th 2007, respectively. The answers obtained were relative to 1408 OSS projects.

We have then matched these data to those retrieved directly from the platform and collected by Notre Dame University (see Madey, 2010). Initially, we dropped 97 projects that were initiated after the first questionnaire in December 2006. Then, we dropped 212 observations, which either lack the necessary information related to the firm (e.g. size, age, and sector) or have missing items from the questionnaire[4]. The combined dataset includes information on 1099 OSS projects out of which 267 of them are being developed in collaboration with a firm. This final sample allows us to analyze the direct effects of firm involvement with specific policy and the coordination by firm as well as the indirect effect of admin's role in the firm as the employee on main duty through interaction with coordination by firm.

We have both projects with and without firms, and this allows us to isolate the effect of the firm presence, provided that we can control for the projects' characteristics both as such, and as a source of firms' selection of projects. To account for these problems and to decrease endogeneity, we thus decided to structure our equations as follows:

1) To assure exogeneity in the regression, we have created three time periods: t_1 spanning the months from August to December 2006, t_2 from August to December 2007, and t_3 from August to December 2008. Stock variables (such as the number of team members) are measured at the end of each period, while flow variables are measured over each whole period.

2) Our dependent variables are measured in period t_3, while our main variables of interest, i.e. policy, coordination styles, are relative to period t_2. We also included the main projects characteristics relative to period in t_2 in order to account for possible biases induced by the projects' intrinsic capabilities.

3) To rule out the selection bias described above we also included projects' characteristics in period t_1, so that we should obtain results net of the initial attractiveness

[4] In order to alleviate the concerns that the results might have been affected from a non-response bias, we run Wilcoxon Mann-Whitney test to compare our variables of interests across sub-samples (sub-sample of observations before and after dropping the observations with missing items). Mean values of both of the dependent variables, File_Size_t_3 and File_Number_t_3, seems statistically significantly different from each other for the two sub-samples. We also compare our main variables of interests across two sub-samples. The results remain unchanged meaning that our results are not caused due to a non-response bias.

projects had for firms and of the degree at which projects may have been open to firms' presence.

3.2. Measures and Validation

We aim to capture the performance of a project on the basis of its productivity. In specific, we use the sum of the total sizes of files (File_Size_t_3) and the number of files generated for each project (File_Number_t_3). These measures enable us to have an idea about the volume of the code developed in the project.

Three sets of items are used to proxy for our variables of interest in order to test the hypotheses put forward (see Appendix A). First set of items relate to firm's direct/indirect involvement in the project and the presence of a policy on OSS. In order to test for Hypothesis 1, we create a binary variable, Direct_Involvement_with_Policy_t_2, which takes the value 1; if the firm is directly involved in the project with a specific policy and 0; otherwise. Hypothesis 2 relates to the management model of the company and it postulates that coordination by profit-oriented firm will a have negative effect on project's performance. We create a binary variable, Coordination_by_Firm_t_2, which takes the value 1; if the firm is the main authority for projects' coordination activities and 0; otherwise. Finally, in order to test for Hypothesis 3, we create a binary variable, Admin/Employee&Main_Duty_t_2, which takes the value ;1 if the admin of the project is an employee of the firm, whose main duty is to work on the project and 0; otherwise.

We control for the remaining categories of the three items from which we created our variables of interest explained above. When testing Hypothesis 1, we control for the binary variables that represent the categories of direct involvement with no specific policy on the project (Direct_Involvement_No_Policy_t_2) and no direct involvement in the project (No_Direct_Involvement_t_2). The latter is the omitted category in the regression analysis. When testing Hypothesis 2, we control for the categories of coordination by formal leadership (Coordination_Formal_Leadership_t_2), coordination by key people through experience (Coordination_Key_People_t_2), informal coordination (Coordination_Informal_t_2), work alone on the project (Coordination_Work_Alone_t_2) and coordination by non-profit organization (Coordination_by_Non_Profit_t_2). We omit the last category in the regression. Finally, when testing Hypothesis 3, we control for the following categories that are derived from the third item; Admin is the owner of the firm (Admin/Owner_t_2), Admin is an employee, who has duties other

than working on the project (Admin/Employee&Not_Main_Duty_t_2) and Admin is not an employee of the firm (Admin/Not_Employee_t_2). We designate the last category to be the baseline category in the regression.

We also introduce two sets of control variables, one of which controls for project characteristics and the other for firm specific characteristics in order to alleviate reverse causality and selection concerns[5]. Control variables related to projects' characteristics that are relative to t_2 include status of the project as declared by the respondents (Proj_Status_Active_t_2, Proj_Status_Respondent_Not_Active_t_2, Proj_Status_Not_Active_t_2), the number of team members (Team_Size_t_2), the degree of the project in the network of projects on sourceforge.net (measured by the average number of other projects participated by the project's team members, Proj_Degree_t_2), the number of messages sent to the project's forums in the time period under analysis (Forum_Msgs_t_2), whether the projects has activated or not advanced tools for code production (i.e. Concurrent Versions System, Coding_Tool_t_2) or for communication (mailing lists and forums, Communication_Tool_t_2). Many more dummy controls at project level have been identified that correspond to the development status of the project (Dev_Status_Mature/Stable_t_2), the function of the software (Topic_Software_Development_t_2), the language spoken by the team members (Language_English_t_2), operating system the software was written for (Operating_System_Linux_t_2, Operating_System_Group_Independent_t_2)[6], programming language the software was written in (Programming_Language_C_C++_Sharp_t_2), the intended audience of the project (Audience_Developers_t_2) and the user interface type (Environment_Api_t_2). In addition, we introduced controls in the selection equation that were only relative to t_1 such as; whether or not the project has declared any trove categories[7] (No_Trove_t_1), project's registration time on Source Forge (Registered_Time_t_1), performance of the project in the same period by using the

[5] The control variables related to project characteristics that are used in the selection equation correspond to t_1 while in the outcome equation we use the same set of controls measured at t_2. There are only a few exceptions that the variable was present for one of the two periods and, thus, was used in the corresponding equation (in selection equation if t_1 and in outcome equation if t_2). For instance, project's status variables correspond to t_2 and are not included in the selection equation. No_Trove_t_2 was dropped due to collinearity. Thus, we only included No_Trove_t_1 in the selection equation.

[6] OS Independent (written in an interpreted language).

[7] SourceForge.net uses Trove system in order to classify projects. The categories include topic, license, intended audience, programming language, intended audience, development status etc. http://sourceforge.net/apps/trac/sourceforge/wiki/Software%20Map%20and%20Trove

number of files created in t_1 (File_Number_t_1) and total sizes of the files created in t_1 (File_Size_t_1). By including these controls, we aim to avoid a potential selection bias that might originate from firms' decision to participate in a project or projects' willingness to receive firms' participation. Firm specific controls, which are relative to t_2, include size, age, sector and region in which the firm is headquartered. Table 1 provides descriptive statistics. Tetrachoric correlations of our main variables of interest are presented in Table 2. There seems to be no problem regarding multicollinearity among these variables. Table 3a and Table 3b present the pairwise correlations among all the variables.

Insert Table 1 and 2 about here

3.3. Methodology

Heckman's two-step procedure, which combines a first stage probit and a second stage OLS model, is used in order to avoid a potential sample selection bias (Heckman, 1979). While the first step corresponds to the estimation of the probit equation by MLE for the sample selection and to the computation of the inverse Mills ratio (Heckman's lambda), the second step corresponds to the estimation of outcome equation on the selected sample by OLS including lambda as an additional regressor. The procedure is also known as 'Heckit model' and it gives consistent estimators.

In our model, the selection equation includes project characteristics at t_1 that may have a critical role in firms' decision to participate actively in the project.[8] In the outcome equation we aim to explore the effect of our main variables of interest, which are relative to t_2, on project's performance measured by number of files created and total sizes of the files created in t_3. We control for firm level characteristics as well as project level characteristics in both equations. All the control variables in the outcome equation are relative to t_2. However, we used the same control variables in the selection equation measured in t_1 to be consistent.

[8] The reduced sample, on which we work on in the selection equation, includes only those projects in which there is firm involvement.

We use the licensing scheme of the software project as the exclusion restriction in the selection equation. Firms may choose the project that they are going to participate based on the licensing terms of the project. Firms, often, try to bundle open source software with complementary proprietary software from which they intend to generate revenues. Licensing terms of the OSS project and its compatibility with the licensing terms of the complementary proprietary product is an important reason why a firm would or would not want to make investment in an OSS project. The licensing term, however, will not have an effect on the amount of code produced. Most people contribute to open source projects with features and capabilities because they want to use the resulting software themselves (Goldman & Gabriel, 2005). Hence, individual users' initial purpose of contribution is not originated from their intentions to redistribute the outcome product but from their intentions to use it for own particular purposes. On the other hand, the main difference among various open source software licenses originates from the conditions of redistribution and not from the terms of use. The individual users, who are interested in using the software and not in redistributing it, will continue contributing to the project no matter what the licensing term under which the initial code has been released. Thus, in theory, the licensing term of the initial code release will not affect how much code generated with the participation of the community.

Insert Table 3a and 3b about here

---

4. Findings

The results of the Heckit model are presented in Table 4. It is possible to observe the direct effects of our main variables of interest- direct involvement with specific policy and coordination by firm- in the first model. In the second model, we introduce the interaction term between employee assignment as administrator and coordination by firm. For each model, the results are presented under two columns one of which corresponds to the outcome equation and the other to the selection equation. We find that firm's direct involvement in the project with specific policy has a positive effect on project's performance as measured by number of files and

sum of sizes of files generated during the project. On the other hand, the coefficient on the binary variable that represents coordination by firm is negative and significant when we use file size as dependent variable. These results are in support of Hypothesis 1 and Hypothesis 2. We observe that the direct effect of the binary variable that identifies the administrator of project as employee of the firm, whose main duty is to work on the project, do not have direct positive effect on project's performance as measured by file size. However, we observe a significant positive effect of Employee/Main_Duty_t_2 variable on performance as measured by number of files generated.

When we introduce the interaction term between this variable and the coordination by firm variable in the second model, we observe a significant coefficient on the aforementioned interaction term. This result suggests that the negative effect of coordination by firm on project's performance is positively moderated if the firm assigns an employee, as the administrator of the project, to work full time on it. Results of the second model highlight the positive effect of presence of a policy to work on the OSS projects. The coefficient on direct involvement (with specific policy) variable maintains its positive and significant effect also after having introduced the interaction term. One may infer that even if the firm is directly involved in the project, without a specific policy on OSS, it will not contribute much to the performance of the project. Looking at different management methods of projects, it is possible to observe that informal coordination has a negative significant effect on project's performance.

Insert Table 4 about here

---

5. Discussion and Conclusion

The findings of this study are three fold. First, being directly involved in a project with a specific policy on OSS has a positive effect on project's performance. Second, coordination by firm has a negative effect on performance. Third, admin as an employee on main duty does not have a direct positive effect on performance. However, it positively moderates the aforementioned relationship.

These findings contribute to knowledge on benefits of co-creation of innovation between firms and communities of users and developers. The role of the firm in a CIP's complex coordination mechanism and its effect on the success of the project is studied within the context of OSS. To our knowledge, there are no studies in the management literature that assesses the effect of having a specific policy, which describes how to work within a community, on the CIP's performance. We also focus on the role of the firm as the main authority in coordinating the CIP and try to assess the effect of it in project's success. As has been stated earlier, there are several means by which projects might be coordinated, such as formal leadership, informal coordination, key people through experience or work alone. Future research might investigate potential effects of these different coordination mechanisms on the benefits garnered from a collaborative mode of innovation.

It would be also interesting to investigate how different participation strategies may affect the success of such collaborative innovation projects. Firms' primary activities on projects vary largely. Some firms participate by writing code, by testing the early versions of the software or by reporting and fixing bugs; some others give logistic and financial support, take part in planning and designing; many others help in diffusing the product by marketing it. How each of these activities contributes to a successful innovation process through which both the firm and community may benefit is a key question to be resolved by future studies.

Along this line, the paper can also directly inform managers on the strategies they should apply to assure long term sustainability of their external knowledge sourcing activities through communities. Managing the boundaries of collaborations is essential. Written rules and guidelines lead to fruitful joint development of software. In light of these findings, one may intuitively think that firm's involvement in a CIP seems to have a positive effect on the success of the project in the presence of formality and written rules. An explanation of such en effect could be that communities benefit from participation of firms due to their managerial capabilities. However, these capabilities should not exert too much control over so that it does not deter participation by volunteer developers. A balance should be struck between the community of developers and the professional firm for sustaining a fruitful collaboration that would bring benefits for both parties. The participative mode of managerial attitude of the administrator of the project, who is an employee of the firm, might be helpful in achieving this

balance. In this way, firm's organization and management may actually fully develop and increase projects' performance.

6. References

Akgün, A. E., Byrne, J., Keskin, H., Lynn, G. S., & Imamoglu, S. Z. (2005). Knowledge networks in new product development projects: a transactive memory perspective. *Information & Management, 42*(8), 1105-1120.

Akgün, A. E., Byrne, J. C., Lynn, G. S., & Keskin, H. (2007). New product development in turbulent environments: Impact of improvisation and unlearning on new product performance. *Journal of Engineering and technology Management, 24*(3), 203-230.

Benkler, Y. (2002). Coase's penguin, or, Linux and "The nature of the firm". *Yale Law Journal*, 369-446.

Bonaccorsi, A., Giannangeli, S., & Rossi, C. (2006). Entry strategies under competing standards: Hybrid business models in the open source software industry. *Management Science, 52*(7), 1085-1098.

Brun, E., Saetre, A. S., & Gjelsvik, M. (2009). Classification of ambiguity in new product development projects. *European Journal of Innovation Management, 12*(1), 62-85.

Burt, R. S. (1992). *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press.

Capra, E., Francalanci, C., Merlo, F., & Rossi-Lamastra, C. (2011). Firms' involvement in open source projects: A trade-off between software structural quality and popularity. *Journal of Systems and Software, 84*(1), 144-161.

Colombo, M. G., Piva, E., & Rossi, L. C. (2010). L'allocazione del potere decisionale nelle imprese open source. Un'analisi empirica. *Economia e Politica Industriale*.

Dahlander, L. (2007). Penguin in a new suit: A tale of how de novo entrants emerged to harness free and open source software communities. *Industrial and Corporate Change, 16*(5), 913-943.

Dahlander, L., Frederiksen, L., & Rullani, F. (2008). Online communities and open innovation. *Industry and innovation, 15*(2), 115-123.

Dahlander, L., & Magnusson, M. G. (2005). Relationships between open source software companies and communities: Observations from Nordic firms. *Research Policy, 34*(4), 481-493.

Dahlander, L., & Wallin, M. W. (2006). A man on the inside: Unlocking communities as complementary assets. *Research Policy, 35*(8), 1243-1259.

Daniel, S. L., Maruping, L. M., Cataldo, M., & Herbsleb, J. D. (2011). *When cultures clash: Participation in open source communities and its implications for organizational commitment*. Paper presented at the ICIS 2011 Proceedings.

Fitzgerald, B. (2006). The transformation of open source software. *Mis Quarterly, 30*(3), 587-598.

Giuri, P., Ploner, M., Rullani, F., & Torrisi, S. (2010). Skills, division of labor and performance in collective inventions: Evidence from open source software. *International Journal of Industrial Organization, 28*(1), 54-68.

Goldman, R., & Gabriel, R. (2005). *Open Source as Business Strategy: Innovation Happens Elsewhere*. San Francisco, CA: Morgan Kaufmann.

Gruber, M., & Henkel, J. (2006). New ventures based on open innovation–an empirical analysis of start-up firms in embedded Linux. *International Journal of Technology Management, 33*(4), 356-372.

Gunz, H., & Gunz, S. (2007). Hired professional to hired gun: An identity theory approach to understanding the ethical behaviour of professionals in non-professional organizations. *Human Relations, 60*(6), 851-887.

Hargadon, A., & Sutton, R. I. (1997). Technology brokering and innovation in a product development firm. *Administrative science quarterly*, 716-749.

Hargrave, T. J., & Van de Ven, A. H. (2006). A collective action model of institutional innovation. *Academy of Management Review, 31*(4), 864-888.

Hars, A., & Ou, S. (2002). Working for free? Motivations of participating in open source projects. *International Journal of Electronic Commerce, 6*(3), 25-39.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society, 47*(1), 153-162.

Henkel, J. (2006). Selective revealing in open innovation processes: The case of embedded Linux. *Research Policy, 35*(7), 953-969.

Henkel, J. (2009). Champions of revealing—the role of open source developers in commercial firms. *Industrial and Corporate Change, 18*(3), 435-471.

Jackson, S. E., Brett, J. F., Sessa, V. I., Cooper, D. M., Julin, J. A., & Peyronnin, K. (1991). Some differences make a difference: Individual dissimilarity and group heterogeneity as correlates of recruitment, promotions, and turnover. *Journal of applied psychology, 76*(5), 675.

Jayaram, J. (2008). Supplier involvement in new product development projects: dimensionality and contingency effects. *International Journal of Production Research, 46*(13), 3717-3735.

Jeppesen, L. B., & Frederiksen, L. (2006). Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *Organization Science, 17*(1), 45-63.

Lakhani, K. R., & Wolf, R. G. (2005). Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. In J. Feller, B. Fitzgerald, S. A. Hissam & K. R. Lakhani (Eds.), *Perspectives on Free and Open Source Software* (pp. 3-22). Cambridge, MA: MIT Press

Laursen, K., & Salter, A. (2004). Searching high and low: what types of firms use universities as a source of innovation? *Research Policy, 33*(8), 1201-1215.

Laursen, K., & Salter, A. (2006). Open for innovation: the role of openness in explaining innovation performance among UK manufacturing firms. *Strategic management journal, 27*(2), 131-150.

Madey, G. (2010). The sourceforge research data archive (SRDA). *University of Notre Dame*.

Mockus, A., Fielding, R. T., & Herbsleb, J. (2000). *A case study of open source software development: The Apache server.* Paper presented at the Software Engineering, 2000. Proceedings of the 2000 International Conference on.

O'Mahony, S., & Bechky, B. A. (2008). Boundary organizations: Enabling collaboration among unexpected allies. *Administrative science quarterly, 53*(3), 422-459.

O'Mahony, S. (2003). Guarding the commons: How community managed software projects protect their work. *Research Policy, 32*(7), 1179-1198.

Oreg, S., & Nov, O. (2008). Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values. *Computers in human behavior, 24*(5), 2055-2073.

Pelled, L. H., Eisenhardt, K. M., & Xin, K. R. (1999). Exploring the black box: An analysis of work group diversity, conflict and performance. *Administrative science quarterly, 44*(1), 1-28.

Roberts, J. A., Hann, I.-H., & Slaughter, S. A. (2006). Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects. *Management Science, 52*(7), 984-999.

Rucht, D. (2004). Movement Allies, Adversaries, and Third Parties. In H. Kriesi (Ed.), *The Blackwell companion to social movements* (pp. 197-216). Malden, MA: Wiley-Blackwell.

Schulze, A., & Hoegl, M. (2006). Knowledge creation in new product development projects. *Journal of Management, 32*(2), 210-236.

Spaeth, S. (2005). *Coordination in Open Source Projects: A Social Network Analysis Using CVS Data.*

Sproull, L., Dutton, W., & Kiesler, S. (2007). Introduction to the special issue: Online communities. *Organization Studies, 28*(3), 277-281.

Stam, W. (2009). When does community participation enhance the performance of open source software companies? *Research Policy, 38*(8), 1288-1299.

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology,translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social studies of science, 19*(3), 387-420.

Stewart, K. J., & Gosain, S. (2006). The impact of ideology on effectiveness in open source software development teams. *Mis Quarterly, 30*(2).

Swink, M., & Song, M. (2007). Effects of marketing-manufacturing integration on new product development time and competitive advantage. *Journal of Operations Management, 25*(1), 203-217.

Tajfel, H., & Turner, J. (1986). *The Social Identity Theory of Intergroup Behavior*. Chicago, IL: Nelson-Hall.

Timmermans, S., & Leiter, V. (2000). The redemption of thalidomide: standardizing the risk of birth defects. *Social studies of science, 30*(1), 41-71.

Trott, P. (1998). *Innovation Management and New Product Development*: Prentice Hall.

von Hippel, E. (1988). The Sources of Innovation. New York: Oxford University Press.

von Hippel, E., & von Krogh, G. (2003). Open source software and the "private-collective" innovation model: Issues for Organization Science. *Organization Science, 14*(2), 209-223.

Wagstrom, P. A., Mockus, A., Herbsleb, J. D., & Kraut, R. (2010). *The impact of commercial organizations on volunteer participation in an online community*. Paper presented at the Academy of Management Conference.

Wasserman, S. (1994). *Social Network Analysis: Methods and Applications* (Vol. 8): Cambridge University Press.

West, J. (2003). How open is open enough?: Melding proprietary and open source platform strategies. *Research Policy, 32*(7), 1259-1285.

West, J., & O'Mahony, S. (2008). The role of participation architecture in growing sponsored open source communities. *Industry and innovation, 15*(2), 145-168.

Table 1: Descriptive Statistics

Variables	*No. Obs*	*Mean*	*Std. Dev.*	*Min*	*Max*
1 File_Size_t_3	267	5.208167	8.259461	0	23.25506
2 File_Number_t_3	267	.623222	1.134998	0	4.795791
3 Direct_Involvement_with_Policy_t_2	267	.3857678	.4876903	0	1
4 Direct_Involvement_No_Policy_t_2	267	.340824	.4748764	0	1
5 Coordination_by_Firm_t_2	1099	.1182894	.3230975	0	1
6 Coordination_Formal_Leadership_t_2	1099	.0500455	.2181381	0	1
7 Coordination_Key_People_t_2	1099	.2183803	.413335	0	1
8 Coordination_Informal_t_2	1099	.2829845	.4506541	0	1
9 Coordination_Work_Alone_t_2	1099	.3066424	.4613095	0	1
10 Admin/Owner_t_2	1099	.0746133	.262886	0	1
11 Admin/Employee&Main_Duty_t_2	1099	.0409463	.1982561	0	1
12 Admin/Employee&Not_Main_Duty_t_2	1099	.0709736	.2568977	0	1
*13 Coord(firm)*Admin/Emp&Main_t_2*	1099	.0209281	.1432089	0	1
14 Firm_Size_t_2	267	3.651685	2.389361	1	8
15 Firm_Age_t_2	267	30.31835	60.11486	4	790
16 Proj_Status_Active_t_2	1099	.055505	.2290676	0	1
17 Proj_Status_Respondent_Not_Active_t_2	1099	.0636943	.2443187	0	1
18 Proj_Status_Not_Active_t_2	1099	.3539581	.4784141	0	1
19 #Messages_on_Forum_t_2	1099	17.96087	180.6526	0	4156
20 Project_Degree_t_2	1099	2.271274	1.660326	1	23.66
21 Dev_Status_Mature/Stable_t_2	1099	.5641492	.4960936	0	1
22 Reg_Time_Every_6_Months_t_2	1099	9.028207	3.817576	1	15
23 Team_Size_t_2	1099	6.007279	9.715814	1	144
24 Topic_Software_Development_t_2	1099	.4131028	.4926152	0	1
25 Prog_Language_C_C++_Sharp_t_2	1099	.2565969	.4369538	0	1
26 Audience_Developers_t_2	1099	.7042766	.456575	0	1
27 Language_English_t_2	1099	.7015469	.4577877	0	1
28 Operating_System_Indep. _t_2	1099	.6797088	.4668008	0	1
29 Operating_System_Linux_t_2	1099	.1856233	.3889794	0	1
30 Environment_Api_t_2	1099	.1656051	.371895	0	1
31 Communication_Tool_t_2	1099	.9399454	.2376961	0	1
32 Coding_Tool_t_2	1099	.7825296	.412713	0	1

Table 2: Correlation Matrix (Tetrachoric Correlations)

	1	2	3	4	5	6	7	8	9	10	11
1 Direct_Involvement_with_Policy_t_2	1.000										
2 Direct_Involvement_No_Policy_t_2	-0.926	1.000									
3 Coordination_by_Firm_t_2	0.351	-0.015	1.000								
4 Coordination_Formal_Leadership_t_2	0.169	-0.155	-0.172	1.000							
5 Coordintion_Key_People_t_2	0.014	-0.208	-0.292	0.246	1.000						
6 Coordination_Informal_t_2	-0.222	0.112	-0.383	-0.456	0.119	1.000					
7 Coordination_Work_Alone_t_2	-0.201	0.164	-0.203	0.047	-0.442	-0.479	1.000				
8 Admin/Owner_of_Firm_t_2	0.194	-0.062	0.144	0.188	-0.014	0.151	-0.079	1.000			
9 Admin/Employee&Main_Duty_t_2	0.032	-0.093	0.085	0.098	-0.008	-0.549	0.005	-0.511	1.000		
10 Admin/Employee&Not_Main_Duty_t_2	-0.058	0.067	-0.082	0.002	-0.056	0.096	0.236	-0.389	-0.514	1.000	
11 Coord(firm)*Admin/Emp&Main_t_2	0.087	-0.005	0.571	-0.442	-0.162	-0.409	-0.084	-0.408	0.735	-0.406	1.000

Table 3a: Correlation Matrix (Pearson Correlations)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 File_Size_t_3	1.00															
2 File_Number_t_3	0.91	1.00														
3 Direct_Involvement_with_Policy_t_2	0.17	0.20	1.00													
4 Direct_Involvement_No_Policy_t_2	-0.14	-0.13	-0.57	1.00												
5 Coordination_by_Firm_t_2	0.01	0.04	0.25	-0.00	1.00											
6 Coordintion_Formal_Leadership_t_2	0.16	0.23	0.09	-0.07	-0.11	1.00										
7 Coordination_Key_People_t_2	0.04	0.02	0.00	-0.13	-0.25	0.11	1.00									
8 Coordination_Informal_t_2	-0.14	-0.15	-0.15	0.09	-0.33	-0.16	0.02	1.00								
9 Coordination_Work_Alone_t_2	-0.08	-0.10	-0.11	0.09	-0.19	-0.05	-0.21	-0.16	1.00							
10 Admin/Owner _t_2	0.03	0.02	0.17	-0.03	0.23	0.06	-0.01	0.01	-0.06	1.00						
11 Admin/Employee&Main_Duty_t_2	0.23	0.29	0.03	-0.05	0.04	0.04	-0.03	-0.20	-0.03	-0.30	1.00					
12 Admin/Employee&Not_Main_Duty _t_2	-0.13	-0.14	-0.02	0.06	0.03	-0.06	-0.04	-0.04	0.14	-0.43	-0.29	1.00				
13 Coord(byfirm)*Adm/Emp&Main_t_2	0.21	0.25	0.06	0.01	0.33	-0.11	-0.08	-0.14	-0.07	-0.20	0.68	-0.20	1.00			
14 Firm_Size_t_2	-0.09	-0.09	-0.14	-0.08	-0.21	-0.10	0.12	0.04	0.01	-0.49	0.16	0.13	0.08	1.00		
15 Firm_Age_t_2	-0.09	-0.08	-0.11	-0.00	-0.11	-0.02	-0.03	0.12	-0.01	-0.20	0.06	-0.01	0.02	0.43	1.00	
16 Proj_Status_Active_t_2	-0.12	-0.10	0.04	-0.04	0.05	-0.03	0.14	0.01	-0.03	0.04	0.04	-0.07	0.00	-0.05	-0.07	1.00
17 Proj_Status_ Respondent_Not_Active_t_2	-0.09	-0.08	0.03	-0.07	-0.04	-0.10	0.05	0.08	-0.05	-0.11	0.08	-0.04	0.02	0.12	-0.01	-0.10
18 Proj_Status_Not_Active_t_2	-0.25	-0.23	-0.01	0.03	-0.11	-0.13	-0.13	0.07	0.03	-0.09	-0.15	0.01	-0.10	0.07	0.12	-0.17
19 # Messages_on_Forum	0.25	0.39	0.11	-0.04	0.05	0.11	0.02	-0.05	-0.05	-0.03	0.10	0.02	0.17	-0.05	-0.03	-0.05
20 Project_Degree_t_2	-0.06	-0.05	0.07	-0.11	-0.06	-0.03	-0.01	0.10	0.14	-0.04	-0.15	0.08	-0.13	-0.07	-0.08	0.01
21 Dev_Status_Mature/Stable	0.16	0.13	0.18	-0.09	0.00	0.00	0.08	-0.06	0.01	0.08	0.12	-0.05	0.11	-0.02	-0.06	0.09
22 Reg_Time_Every_6_Months_t_2	-0.04	-0.06	-0.03	0.12	0.13	0.05	-0.17	0.10	-0.07	0.07	-0.15	0.04	0.02	-0.10	-0.02	-0.19
23 Team_Size_t_2	0.43	0.54	0.09	-0.11	-0.04	0.23	0.22	-0.10	-0.16	-0.12	0.29	-0.07	0.18	-0.01	-0.02	-0.03
24 Topic_Software_Development_t_2	-0.03	-0.03	-0.02	0.05	-0.04	-0.01	0.16	0.08	-0.04	0.04	-0.02	-0.00	0.10	0.01	-0.09	0.00
25 Prog_Language_C_C++_Sharp_t_2	-0.01	0.01	-0.02	-0.03	-0.06	0.02	0.04	-0.06	0.07	-0.01	0.05	-0.08	-0.08	0.05	-0.02	0.08
26 Audience_Developers_t_2	-0.04	-0.01	-0.04	0.03	0.04	-0.01	0.14	0.02	0.06	0.02	0.07	0.05	0.08	0.01	-0.10	-0.01
27 Language_English_t_2	0.06	0.10	0.11	-0.09	0.01	0.03	0.06	-0.09	-0.04	-0.05	0.11	-0.04	0.04	-0.04	-0.12	0.09
28 Operating_System_Indep._t_2	0.00	0.00	-0.03	-0.03	0.08	-0.04	0.01	-0.03	0.02	0.00	-0.05	0.08	0.02	-0.09	-0.02	0.10
29 Operating_System_Linux_t_2	-0.03	-0.05	0.03	0.03	0.01	0.11	-0.09	-0.16	0.07	0.01	-0.04	0.02	-0.08	0.03	0.02	-0.01
30 Environment_Api_t_2	0.08	0.05	0.07	-0.01	0.19	0.06	-0.02	-0.13	-0.11	0.15	0.03	-0.12	0.02	-0.16	-0.09	-0.03
31 Communication_Tool_t_2	0.02	0.04	-0.04	0.12	-0.08	0.10	-0.06	-0.03	0.06	0.04	-0.06	-0.01	-0.07	-0.05	-0.01	-0.23
32 Coding_Tool_t_2	-0.30	-0.27	-0.06	0.05	-0.04	-0.00	-0.08	-0.09	0.11	-0.04	-0.04	0.03	-0.12	0.08	0.10	-0.03

Table 3b: Correlation Matrix (Pearson Correlations)

	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
17 Proj_Status_ Respondent_Not_Active_t_2	1.00															
18 Proj_Status_Not_Active_t_2	-0.12	1.00														
19 # Messages_on_Forum	-0.03	-0.06	1.00													
20 Project_Degree_t_2	0.16	-0.09	-0.05	1.00												
21 Dev_Status_Mature/Stable	0.01	-0.15	0.09	-0.03	1.00											
22 Reg_Time_Every_6_Months_t_2	0.03	-0.08	0.09	-0.12	-0.23	1.00										
23 Team_Size_t_2	-0.06	-0.12	0.46	-0.08	0.17	-0.23	1.00									
24 Topic_Software_Development_t_2	0.11	-0.05	-0.00	-0.09	-0.03	-0.08	-0.03	1.00								
25 Prog_Language_C_C++_Sharp_t_2	-0.00	0.04	-0.06	0.16	-0.04	-0.39	0.05	-0.04	1.00							
26 Audience_Developers_t_2	0.09	-0.10	0.03	0.09	0.10	-0.14	0.08	0.41	0.10	1.00						
27 Language_English_t_2	-0.01	0.08	0.08	0.05	0.19	-0.44	0.11	-0.02	0.20	0.03	1.00					
28 Operating_System_Indep._t_2	-0.07	-0.06	-0.04	0.00	0.06	-0.09	0.01	0.10	-0.14	0.06	0.17	1.00				
29 Operating_System_Linux_t_2	-0.02	0.08	-0.02	0.06	0.06	-0.20	-0.05	-0.04	0.29	-0.07	0.11	-0.25	1.00			
30 Environment_Api_t_2	0.03	-0.09	-0.04	-0.09	0.06	0.24	-0.01	-0.08	-0.10	-0.04	0.08	0.03	0.04	1.00		
31 Communication_Tool_t_2	-0.04	0.06	0.04	-0.04	-0.06	0.11	0.09	-0.04	0.05	0.04	-0.01	-0.08	0.03	0.06	1.00	
32 Coding_Tool_t_2	0.10	0.20	-0.12	0.07	-0.11	-0.17	-0.13	-0.09	0.17	0.01	0.07	0.00	0.15	-0.03	0.22	1.00

Table 4: Results of the Heckman Model

	Model I		Model II		Model I		Model II	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Variables	***File_Size_t₃***	***Selection***	***File_Size_t₃***	***Selection***	***File_No_t₃***	***Selection***	***File_No_t₃***	***Selection***
Firm involvement in the project								
H1 *Direct_Involvement_with_Policy_t₂*	2.271*		2.246*		0.507***		0.507***	
	(1.168)		(1.176)		(0.143)		(0.149)	
Direct_Involvement_No_Policy_t₂	-0.064		-0.104		0.163		0.165	
	(1.180)		(1.181)		(0.129)		(0.130)	
Coordination style								
H2 By_Firm_t₂	-2.002*		-3.323**		-0.210		-0.371**	
	(1.128)		(1.338)		(0.141)		(0.155)	
Formal_Leadership_t₂	1.279		1.893		0.484**		0.603**	
	(1.704)		(1.766)		(0.229)		(0.258)	
Key_People_Through_Experience_t₂	-0.961		-1.257		-0.277**		-0.316***	
	(1.013)		(1.022)		(0.114)		(0.114)	
Informal_Coordination_t₂	-2.447*		-2.998**		-0.222		-0.291*	
	(1.372)		(1.418)		(0.154)		(0.157)	
Work_Alone_t₂	-1.383		-1.817		-0.156		-0.223	
	(1.303)		(1.287)		(0.154)		(0.155)	
Admin involvement in the firm								
Admin/Employee&Main_Duty_t₂	1.516		-1.352		0.388**		0.030	
	(1.436)		(1.723)		(0.176)		(0.210)	
Admin/Employee&Not_Main_Duty_t₂	-2.010*		-1.562		-0.140		-0.075	
	(1.166)		(1.174)		(0.130)		(0.147)	
Admin/Owner_t₂	-0.216		0.297		0.142		0.206	
	(1.273)		(1.260)		(0.155)		(0.159)	
Interaction term								
H3 *Admin/Employee&Main_Duty_t₂ * Coord_By_Firm_t₂*			5.677**				0.739**	
			(2.388)				(0.365)	
Controls for firm-level characteristics								
Firm_Size_t₂	-0.106		-0.110		0.002		0.001	
	(0.204)		(0.203)		(0.026)		(0.030)	
Firm_Age_t₂	-0.007		-0.006		-0.001		-0.001	
	(0.006)		(0.006)		(0.001)		(0.001)	
Sector_Dummies_t₂	Included		Included		Included		Included	
Region_Dummies_t₂	Included		Included		Included		Included	
Controls for project characteristics								
Proj_Status_Active _t₂	-3.812***		-3.651***		-0.345**		-0.329**	
	(1.057)		(1.015)		(0.139)		(0.141)	
Proj_Status_Respondent_Not_Active_t₂	-3.070**		-2.678*		-0.380*		-0.331	
	(1.536)		(1.564)		(0.209)		(0.236)	
Proj_Status_Not_Active_t₂	-3.885***		-4.041***		-0.326***		-0.334***	
	(0.971)		(0.979)		(0.121)		(0.121)	
Communication_Tool_(t₂/t₁)	3.080*	-0.225*	2.793	-0.231*	0.516**	-0.206	0.477**	-0.218*
	(1.711)	(0.135)	(1.742)	(0.135)	(0.220)	(0.126)	(0.223)	(0.128)
Coding_Tool_(t₂/t₁)	-3.064***	-0.284***	-2.970**	-0.284***	-0.174	-0.225**	-0.159	-0.222**
	(1.187)	(0.085)	(1.178)	(0.085)	(0.173)	(0.088)	(0.195)	(0.089)
Team_Size_(t₂/t₁)	0.140*	0.011***	0.138*	0.011***	0.039***	0.009***	0.037***	0.009***
	(0.076)	(0.002)	(0.074)	(0.002)	(0.010)	(0.003)	(0.010)	(0.002)
Dev_Status_Mature/Stable_(t₂/t₁)	-0.650	0.154**	-0.877	0.151**	-0.192	0.130*	-0.216*	0.124*
	(0.963)	(0.075)	(0.950)	(0.074)	(0.127)	(0.067)	(0.131)	(0.066)
Topic_Software_Development_(t₂/t₁)	-1.390	0.213***	-1.844*	0.212***	-0.205	0.218***	-0.261*	0.220***
	(1.028)	(0.080)	(1.058)	(0.080)	(0.141)	(0.079)	(0.144)	(0.079)

Language_English_(t_2/t_1)	-0.519	0.043	-0.450	0.043	0.075	0.012	0.069	(0.098)
	(1.269)	(0.082)	(1.248)	(0.082)	(0.170)	(0.087)	(0.168)	0.008
Operating_System_Linux_(t_2/t_1)	-0.620	0.045	-0.690	0.042	-0.206	-0.009	-0.223	-0.023
	(1.406)	(0.110)	(1.385)	(0.110)	(0.177)	(0.134)	(0.178)	(0.166)
Operating_System_Group_Indep. _(t_2/t_1)	-0.172	0.106	-0.261	0.108	-0.038	0.044	-0.045	0.046
	(1.146)	(0.085)	(1.127)	(0.085)	(0.164)	(0.085)	(0.162)	(0.085)
Programming_Language_C_C++_Sharp_(t_2/t_1)	-0.352	0.226**	-0.194	0.215**	-0.021	0.226**	0.010	0.217**
	(1.103)	(0.097)	(1.089)	(0.097)	(0.142)	(0.088)	(0.142)	(0.088)
Environment_Api_(t_2/t_1)	0.069	0.237**	0.406	0.230**	-0.124	0.261***	-0.066	0.247**
	(1.208)	(0.100)	(1.226)	(0.100)	(0.153)	(0.094)	(0.153)	(0.108)
Audience_Developers_(t_2/t_1)	-2.982**	0.233**	-2.803*	0.228**	-0.383*	0.206**	-0.371*	0.190*
	(1.491)	(0.096)	(1.467)	(0.096)	(0.202)	(0.093)	(0.222)	(0.098)
Registered_Time	-0.430**	0.044***	-0.480***	0.044***	-0.049**	0.036***	-0.058**	0.036***
	(0.176)	(0.012)	(0.176)	(0.012)	(0.025)	(0.011)	(0.027)	(0.011)
File_No_t_1						0.274***		0.279***
						(0.037)		(0.040)
File_Size_t_1		0.026***		0.027***				
		(0.006)		(0.005)				
No_Trove_t_1		1.907***		1.920***		1.599***		1.560***
		(0.452)		(0.440)		(0.328)		(0.354)
Exclusion Restrictions								
License_Apache_t_1		-0.153		-0.131		-0.049		-0.032
		(0.172)		(0.169)		(0.149)		(0.153)
License_Artistic_t_1		-0.311		-0.327		-0.240		-0.247
		(0.314)		(0.312)		(0.306)		(0.295)
License_BSD_t_1		0.194*		0.194*		0.216**		0.209*
		(0.118)		(0.114)		(0.106)		(0.115)
License_GPL_t_1		-0.318***		-0.305***		-0.265**		-0.252**
		(0.103)		(0.103)		(0.103)		(0.120)
License_LGPL_t_1		0.084		0.105		0.132		0.164
		(0.099)		(0.098)		(0.090)		(0.107)
License_MIT_t_1		-0.066		-0.030		0.036		0.079
		(0.241)		(0.235)		(0.255)		(0.260)
License_Other_t_1		-0.008		0.002		0.018		0.035
		(0.101)		(0.100)		(0.087)		(0.086)
License_Prop_Other_t_1		0.348**		0.372**		0.328*		0.369*
		(0.154)		(0.156)		(0.187)		(0.200)
License_Pub_Domain_t_1		-0.025		-0.010		-0.133		-0.111
		(0.299)		(0.295)		(0.346)		(0.373)
Constant	23.053***	-1.306***	24.774***	-1.311***	2.313***	-1.271***	2.580***	-1.263***
	(4.301)	(0.260)	(4.395)	(0.261)	(0.760)	(0.237)	(0.981)	(0.239)
Observations	1,099	1,099	1,099	1,099	1,099	1,099	1,099	1,099

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1