Paper to be presented at the DRUID Summer Conference 2007

on

APPROPRIABILITY, PROXIMITY, ROUTINES AND INNOVATION

Copenhagen, CBS, Denmark, June 18 - 20, 2007

A TEST OF LAZEAR'S THEORY OF ENTREPRENEURSHIP IN THE OPEN SOURCE SOFTWARE VIRTUAL COMMUNITY

Paola Giuri
LEM - Sant'Anna School of Adavanced Studies
giuri@sssup.it

Francesco Rullani
IVS - Copenhagen Business School
fr.ivs@cbs.dk

***Salvatore Torrisi**
Department of Management, University of Bologna
torrisi@unibo.it

Abstract:
This paper studies the emergence of entrepreneurs and their skill profile in the open source software community. We test the hypothesis that entrepreneurs, carrying out complex, multi-task activities, have more balanced skill sets compared with individuals who are less involved in project management activities.
Our empirical analysis employs the SourceForge dataset containing information on 77,039 individuals working in 54,229 OSS projects. We estimate logit and ordered logit models to explore the likelihood that an individual is a project founder or manager. Our main regressors include individual attributes like skill level and diversity, and project-level controls. Results support our hypothesis.

JEL - codes: O31, O32, -

A test of Lazear's theory of entrepreneurship in the Open Source Software virtual community

Summary

This paper studies the emergence of entrepreneurs and their skill profile in the open source software (OSS) community. We test the hypothesis that entrepreneurs, carrying out complex, multi-task activities, have more balanced skill sets compared with individuals who are less involved in project management activities.

Our empirical analysis employs the SourceForge dataset containing information on 77,039 individuals working in 54,229 OSS projects. We estimate logit and ordered logit models to explore the likelihood that an individual is a project founder or manager. Our main regressors include individual attributes like skill level and diversity, and project-level controls. Results support our hypothesis.

Keywords: entrepreneurship, skills, managers, founders, open source software, virtual community

1. Introduction

According to Schumpeter the entrepreneurial function “does not necessarily consist in either inventing anything or otherwise creating the conditions which the enterprise exploits. It consists in getting things done” (Schumpeter, 1942, p.132). From this perspective, entrepreneurship does not consist exclusively in the creation of a new company or business but also in the management and coordination of innovative combinations of human capital and other inputs with the purpose of ‘getting things done’.

Our analysis draws on this broad view of entrepreneurship to look at the emergence of entrepreneurs and their skill profile in the open source software (OSS) community. In this particular context we define entrepreneurs as both founders and managers of open source projects.

Various empirical accounts of the OSS community suggests that many projects rely on a quite well defined division of labor between ‘core developers’ and project leaders, who control the evolution of the base code, and a wider ‘periphery’ of contributors who provide feedbacks that are critical for product quality improvement (the “obscure developers” as defined by Dalle and Jullien, 2003).

In large projects like Linux or Apache there is a quite clear decision making chain. For example, Linus Torvalds, the founder of the Linus project, has the right to decide which changes enter the official Linux releases while core developers of the Apache Group approve changes to the source code by voting (Mockus et al., 2000). Even in these large projects, however, developers contribute on a voluntary basis, choosing which tasks to undertake and how much effort to devote. Project founders and leaders then must be able to manage a virtual team in a regime of 'distributed authority' (Mateos Garcia and Steinmueller, 2003).

Several OSS projects are participated by a small number of active, highly committed programmers (Ghosh and David, 2003). Most probably, in small projects the division of labor is more blurred than in large ones and the coordination efforts of project leaders are quite limited. But this does not imply that these projects do not need any project management. On the contrary, the lack of management and coordination is probably a major source of failure of many OSS projects. In this setting, it is crucial to understand the forces that lead to the emergence of project managers among OSS participants.

The knowledge of these forces is still limited and we aim to fill this gap in the literature by analyzing the association between the roles played by individuals in the projects with their skill profile. In line with the theory of occupational choice elaborated by Lazear (2002, 2004), we expect that OSS entrepreneurs must possess a diversified skill set that is needed to select the inputs provided by various participants, motivate contributors and coordinate their efforts while specialists, like pure developers, are supposed to be endowed with a more focused skill set.

We test this hypothesis by controlling for various individual and project characteristics such as project size and the degree of modularity of the development process. Modular, flexible design and manufacturing systems are based on multi-skilled teams and have to be coordinated by novel management approaches that differ from traditional ones. Spatial dispersion of team members and computer-mediated interaction which are typical of 'virtual teams' pose additional challenges to project management and requires individuals focusing on this type of activity (Martins et al., 2004). Project team managers must be 'able to articulate project goals and to assign responsibilities with specific schedules and work deadlines. They have to be flexible and willing to delegate responsibilities to other group members and help developing links among participants thus favoring mutual trust' (Kayworth and Leidner, 2000: 189).

Our empirical analysis employs a dataset containing information on 77,039 individuals working in 54,229 projects hosted by the SourceForge.net website from November 1999 to January 2003.

The SF.net dataset provides information on the founder of each project and on the role of each user registered with a project (i.e., developer, project manager, web designer, content manager, 'all hands persons', etc.). For our purposes, we grouped these roles into broader role categories implying increasing levels of managerial task complexity.

We explore the likelihood that an individual plays multi-task, managerial roles by estimating an ordered logit model. Our main regressors include individual attributes such as skill level and diversity, the average project size and the average number of subprojects per project as a proxy for project modularity, plus a series of controls. We also perform logit estimations of the probability to be a founder of the project, conditional to individual level characteristics.

The results support the hypothesis that entrepreneurs, carrying out complex, multi-task activities, have more balanced skill sets compared with individuals who are less involved in project management activities.

The paper is organized as follows. Section 2 discusses the background literature and the research hypothesis. Section 3 illustrates the data used in the empirical analysis, descriptive statistics and the econometric estimations. Section 4 concludes.

2. Theory and research hypotheses

The OSS context

The organization of OSS projects can benefit from recent advances in the analysis of 'virtual teams'. These have been defined as 'teams whose members use technology to varying degrees in working across locational, temporal, and relational boundaries to accomplish an interdependent task' (Martins et al., 2004, p. 808). This stream of the literature recognizes that virtual teamwork has become widespread in several activities, especially in knowledge-intensive industries. For instance, Kanawattachai and Yoo (2002) show that over 60% of professional employees work in virtual teams. But, despite the growing number of papers produced in over ten years, 'there is a lack of clarity on what we know' on virtual teams (Martins et al. 2004). An area that has attracted limited research efforts so far (e.g., Kayworth and Leidner, 2000) is about leadership in virtual environments. Drawing on a large survey of the management literature, Martins et al. (2004) note that relevant for future research are issues such as the way leaders define roles, motivate and evaluate participants' performance, and how the degree of virtualness affects the patterns of leader-member exchange (p. 821).

OSS represents an ideal testing ground for analyzing virtual teams and indeed several papers have addressed their attention to the organization of OSS projects and the motivations of participants (see, for instance, Harhoff et al., 2003; von Hippel, 2001; Lerner and Tirole, 2002).

The OSS model of innovation involves a large community that shares values (generalized reciprocity and meritocracy), conduct rules and institutions (such as priority, peer review principles, and the refuse of formal development methodologies and management systems).

These social norms and institutions conform to the concepts of self-organization and gift economy rather than the principles of engineering and market economy (e.g., Raymond, 1999, 2001; Di Bona et al., 1999).

The increasingly large number of programmers who offer their voluntary contribution to OSS projects has spurred a stream of research which focuses primarily on specific OSS projects, especially large ones. For instance, Koch and Schneider (2002) have analyzed the CVS (concurrent versions system) of the GNOME project (an OSS project dedicated to a desktop environment for users and an application framework for developers) and found that only a small number of programmers work together on the same file. The number of co-developers increases with the size of the file and more active programmers work more for large files compared with less active programmers. Similarly, Mockus et al. (2000) found that in the case of the Apache server project the top 15 developers contributed more than 83% of changes to Apache source code and only 25 developers submitted changes on a regular basis during the period 1995-1999. The 15 most active 'core developers' corresponds approximately to the Apache Group, the organization responsible for the management of the Project. These project maintainers are primarily devoted to developing or reviewing new functionalities to the base code and, to a lesser extent, to fixing defects. The periphery of less active contributors is made of non core developers (about 250 people during the time window of Mockus et al.'s analysis) which, relative to core developers, are more active in bugs or problem-related changes (patches). The most external part of the Apache server's periphery is made of a wider community of over 3,000 users who only report bugs (Mockus et al., 2000).

We should remember, however, that the majority of OSS projects are small and much less organized than that described above. For example, Krishnamurthy (2002) has analyzed the top 100 mature projects in Sourceforge and found that the median number of co-developers was only 4. These projects then generate only limited informal exchange among users and therefore require less coordination. But the leaders of small projects most probably carry out different activities and have to work hard to attract new contributors. To some extent then project leaders represent a critical resource for both large and small OSS projects.

It is worth to note that coordinating and managing an OSS project appears to be a particularly challenging task compared with other settings like a traditional firm. As Lerner and Tirole (2002) pointed out, OSS leaders have to carry out some critical tasks: a) to provide a 'vision' that is provided through a critical mass of code that demonstrates her expertise and credibility; b) to attract new programmers by posing challenging issues and, at the same time, leaving to potential contributors significant opportunities for future improvements to the initial code; c) to ensure an efficient division of the project into modules and to allow contributors to perform their tasks independently from the rest of the contributors; d) to avoid that conflicting views and approaches among participants lead to dropouts and forking (p. 21).

When this critical role materializes in OSS projects? We submit to empirical test the view that the emergence of entrepreneurs in OSS teams is associated with the skill profile of team members. In our empirical analysis we also control for characteristics of the project like size and modularity.

Lazear's theory of entrepreneurship

The economics and strategic management literature of the 1980s and the 1990s has examined different ways in which skill level and diversity may affect firm performance.

First, skill heterogeneity implies that firms can experiment complex combinations of skills that are difficult to imitate (Lippman and Rumelt, 1982). Second, skill diversity allows a more flexible strategic adaptation to changing external environment (Galunic and Rodan, 1998). Skill heterogeneity provides firms with more comprehensive problem solving ability and creative conflict resolution (Sutton and Hargadon, 1997; Galunic and Rodan, 1998). The cognitive diversity resulting from interaction among people with different perspectives makes it possible to identify and formulate a wider array of problems and to find a larger set of alternative solutions (Bantel and Jackson, 1989)[1].

More recently the literature has stressed that the level of capabilities and skills is also particularly important for newly established firms. Well balanced founding teams (or highly skilled single founders) are able to attract financial resources, customers and collaborators (see, for instance, Bhidé, 2000; Baron and Hannah, 2002). In the case of OSS, skilled core developers are more likely to attract new users because their software addresses relevant problems that are not met by commercial products or because it raises technical puzzles that are challenging to the community of developers. This in turn helps the project to evolve and become more productive. OSS projects then are not very different from the traditional entrepreneurial sector, where new ventures have to overcome the 'liability of newness' and

---

[1] Empirical papers on the benefits of heterogeneous workforce mainly provide evidence of positive or curvilinear (inverted U-shape) relations between skill diversity and performance. See Bantel and Jackson (1989), Hamilton et al. (2003), Laursen et al. (2005), Carbonell and Rodriguez (2006). See also Hambrick et al. (1996) on the impact of heterogeneity of top management teams on firms competitive actions.

must convince potential stakeholders to pool their resources to support new ideas and to grow.

Our main hypothesis is that individuals who choose to found a project and/or to become project leaders (either deliberately or as the result of internal 'myopic' selection process) must possess a balanced skill set. They are not necessarily the most skilled among project members neither they need to have a comparative advantage in specific skills. Our hypothesis draws on the Lazear's theory of entrepreneurship (Lazear, 2002, 2004).

Unlike other theories of entrepreneurship which view entrepreneurs as technical specialists, in the Lazear's model of occupational choice entrepreneurs are 'generalists', i.e., multi-skilled individuals. Using a data set of Stanford Master of Business Administration alumni, Lazear finds supports to the 'jack-for-all-trades' theory. In line with the entrepreneurship literature, Lazear (2002) defines entrepreneurs as those individuals who have established the business and 'are usually responsible for the conception of the basic product, hiring the initial team, and obtaining at least some early financing' (p. 3).

Other scholars have also found support to this theory drawing different data sets. For example, Wagner (2003, 2006), using data on the German population and a survey of new entrepreneurs, and Astebro (2006), drawing on data from Canada, have found that the likelihood of being an entrepreneur is higher for individuals with more balanced skill and working experience.

One may wonder why we need a theory of entrepreneurship to explain OSS project founders and leaders and whether OSS entrepreneurs are really the equivalent of entrepreneurs in business enterprises. We believe that OSS founders and project leaders share some important characteristics with entrepreneurs. First, as we discuss later, in most cases project managers are also founders of OSS projects. Moreover, they play a leading role, by coordinating the efforts of different members and finalizing the results of the development team. In fact, Lazear (2002) tested his model by using two different definitions of entrepreneur. The first definition described before refers to new business founders, while the second one includes individuals who take responsibility for the organizations' direction, or major business function. This category includes senior or high-level managers. Lazear finds that these individuals too have a balanced skill set and therefore the 'jack-for-all-trades' view can be applied not only to founders of new initiatives but also to individuals who occupy high-level managerial positions within the organization.

3. Empirical analysis

3.1. Data

Our empirical analysis uses a rich dataset containing information on the role, skill profiles and activities of individuals registered with an OSS project. The dataset has been built from data provided by SourceForge.net (http://sourceforge.net, *SF.net* henceforth) from November 1999 to January 2003. *SF.net* also provides complementary data on several characteristics of the project participated by the individual members.

Our version of the dataset consists of 65,535 projects and 544,669 individuals. These individuals include both people registered with one or more projects (i.e., participants who are supposed to be 'active' contributors) and individuals who registered with the SF.net website but have not become member of any specific project, i.e., users who download software and signal bugs or post questions to the project forums. The total number of people registered with a project (project members) in our dataset is 90,255. Of these, 26,314 provide data on skills. Our sample excludes all 'inactive' projects, i.e. projects with no registered users as to January 2003 or not labeled as "active" by the SF.net staff.

We end up with a sample of 77,039 individuals who registered with 54,229 projects. Data on skills are available for only 23,560 individuals registered with 26,254 projects. Information about project founders is available for 66,944 users.

A main limitation of the SF.net dataset is that it excludes some large OSS projects like Apache and Linux, although other large and popular projects are included. Moreover, a large number of projects are inactive (Howison and Crowston, 2004). To account for these drawbacks we control for the size of projects and level of activity of individual and projects.

3.2.Definition of variables

3.2.1. Dependent variables: the roles of project members

The main contribution of this paper is about the skill profiles of founders and project members playing managerial roles. The SF.net dataset specifies the role of the member registered with each project at the end of the covered period, i.e. January 2003. Since individuals can be registered with more than one project, they can also perform different roles in different projects. Over 22 per cent of individuals in our sample are registered with more than one project (the average number of projects participated by the sample individual is 1.39). Matching individuals and projects yields 106,823 individual-roles pairs. Obviously, the number of pairs is larger than that of individuals (77,039) since, as mentioned before, an individual may perform different roles in different projects of which he/she is member. Table A1 in the Appendix reports the distribution of all pairs by type of role.

Developer is the most frequent role (28.09% of cases), followed by Project manager (10.57%). Other roles like all-hands person, web designer, tester, graphic, are much less frequent while in about half cases project members are not assigned any specific role.

For our purposes, we grouped the 19 roles listed in Table A1 into four categories characterized by increasing levels of managerial task complexity (variable ROLE4). At the lower end of the complexity range there is obviously “No role”. In all likelihood individuals in this category have a very limited commitment in a specific project. The category “Other roles” (Other) includes secondary, supporting tasks like web design, test, editorial/content writer, and consultant. The category “Developer” (DEV), the most frequent across project members, includes members who carry out core project activities which may imply high levels of technical sophistication but limited managerial complexity. The category Project Manager (PM) represents, by definition, the tasks with the highest level of managerial complexity. Project managers may be responsible for specific project modules and subprojects or the whole project.

Table 1 summarizes the distribution of the four categories of roles in our dataset. The first column shows the frequency of each role amongst the individuals in our sample.

[Table 1 around here]

As each individual may perform different roles in different projects we want to see the level of managerial complexity of tasks carried out by individuals across all projects of which they are members. To this end, we consider all roles each individual performs in all projects he/she participates and build a second classification which differs from ROLE4 for one category that represents individuals who play the role of project manager in one project and other roles in other projects. More precisely, the fifth category includes “PM&Other”, “PM&Dev” and “PM&Other&DEV”.

Table 2 shows the distribution of individuals across the five categories above (ROLE5). The share of individuals with No role (44.59%) is lower than in the previous classification suggesting that most individuals registered with more than one project focus their activities on one project in particular while play no specific role in other projects. Individuals with Other roles are 10.24% of the sample. Developers are again the most frequent category of project members (31.91%). Project managers account for 10.49% of individuals while multi-role individuals (category 5) account for 2.78% of the sample individuals.

We use this ordered classification of member roles to estimate the association between several individual and project-level characteristics and the probability that a participant becomes involved in activities that imply increasing levels of managerial complexity.

The results discussed in this paper focus on individuals as unit of analysis and therefore ROLE5 is used as dependent variable.

[Table 2 around here]

From SF.net we also built, for a sample of 66,944 users, a dummy variable called FOUNDER which is equal to 1 if an individual founded at least one SF.net project and equal to 0 otherwise. The last column of Table 2 shows the share of individuals who founded a project in each role category. It is worth noting that 70% of project managers and 82% of individuals who perform multiple roles (project managers and other roles) are also project founders. By contrast, only 24% of developers and 30% of individuals with other roles founded at least a project. Surprisingly, about 82% of people with no role are a project founder. However, we found that on average these individuals (and their projects) have an extremely low level of activity (bugs, patched, messages, etc.) compared with individuals with a specified role. These individuals represent the large number of OSS participants who enter the community by founding a project and remain inactive thereafter. Thus, they are very different from individuals who play a specific role in active projects.

The variable FOUNDER is used as dependent variable in logit estimations of the probability to found an OSS project associated to several individual level characteristics.

3.2.2. Key regressors: skill level and diversity of individuals

A key variable in our analysis is represented by skills. At the time of registration with SF.net, the website users are asked to self-assess their experience in 33 types of different skills which can be grouped into three main areas of expertise: programming languages (e.g., C/C++ and Python), application-specific skills (e.g., networking, security, etc.) and 'people' skills proxied by the knowledge of spoken languages. Registered individuals can also update the information relative to their skills at any time. Unfortunately, the dataset does not tell whether and when registered users have updated their skills since registration and this gives rise to problems of endogeneity. Although the mix and the level of skills declared at the time of registration may change over time, skills are likely to change slowly over time relative to roles and this moderates the endogeneity of this variable.

Another possible drawback of these measures is that they might be affected by self-assessment biases. However, we believe that this is not a serious problem in our case because the information supplied by developers can be made public to other developers who can check its reliability. Since, as noted by Lerner and Tirole (2002), expected delayed benefits arising from signaling represent an important incentive to contribute to the OSS community, individuals who register with SF.net have strong reasons to provide information as close as possible to the reality. Moreover, there are no reasons to believe that a potential bias due to self-reporting should affect particular types of developers or projects.

Drawing on this information, we build the following measures of skills at the individual level.

EXPERIENCE. This variable is computed as the average level of experience of the skills of each individual. The level of experience in each skill is measured on a five point Likert scale, i.e. 1 = less than 6 months; 2 = 6 months-2 years; 3 = 2-5 years; 4 = 5-10 years; 5 = more than 10 years.

N_SKILLS. This variable indicates the number of different skills mastered by the individual member. It is a proxy for the variety of skills of the individual.

HERF. This variable measures the degree of skill diversification of the project member. To build this measure, we first define the share of each skill *i* on the total skills experience of the individual *j*:

$$s_{ij} = S_{ij} * e_{ij} / \sum_{i=1}^{n} S_{ij} * e_{ij}$$

where S_{ij} = 1 when individual *j* reports skill *i*, and 0 otherwise, and e_{ij} represents the experience of the skill *i* of the individual *j* (see EXPERIENCE above). We then computed the Herfindahl index of skills for each individual as follows: $Herf_j = \sum_{i=1}^{n} s_{ij}^2$.

This measure allows weighting each skill of the individual by its experience in each of the skills. For a more intuitive interpretation of the variables based on the Herfindahl indexes in our analysis we use the 1-HERF. The index ranges from 0 (min. skill diversification) to 1 (max skill diversification).

D_MISSING_SKILLS. Dummy equal to 1 if the individual does not report its skills at the time of registration at *SF.net*.

D_VIEW_SKILLS. Dummy variable equal to 1 if the individual agrees that his/her skills can be published in the SF.net website.

3.2.3. Individual-level controls

TIME_REG. Number of months the individual has been registered with SF.net.

D_MAIL. Dummy equal to 1 if the email address of the individual is reported in the dataset.

D_MAIL_COM. Dummy equal to 1 if the e-mail address of the individual has a .com suffix, considered as a proxy for the affiliation to a commercial organization.

N_MESSAGES. Number of messages posted by the individual to project forums hosted by SF.net from November 1999 to January 2003.

N_MAINCONTR_SUB. Number of main contributions (bugs, patches, and feature requests) submitted by the individual to SF.net projects from November 1999 to January 2003.

N_MAINCONTR_ASS. Number of main contributions (bugs, patches, and feature requests) assigned to the individual by project administrators from November 1999 to January 2003.

N_OTHERCONTR_SUB. Number of other contributions (like support requests) submitted by the individual to SF.net projects from November 1999 to January 2003.

N_OTHERCONTR_ASS. Number of other contributions (like support requests) assigned to the individual by SF.net project administrators from November 1999 to January 2003.

D_NL. 13 dummy variables for the natural language spoken by the individual.

3.2.4. Project-level controls

SIZE. An important factor that can explain the emergence of managerial roles is the size and internal organization of the project. In order to estimate the net effect of the skill profiles on the probability to assume a managerial role we need to control for these project characteristics in our econometric analysis. These variables are instead not included in the estimation of the probability to become a project founder, which is obviously independent of ex-post characteristics of the project.

We expect that the likelihood that a project member plays managerial roles is positively affected by the size of the project. This is because with the increase in the number of participants coordination and integration of different inputs become more complex tasks which call for expert project managers.

Our measure of project size is the number of members registered with the project in which the individual is registered at January 10, 2003. If the individual is registered to more than one project, this is the average size of the projects at which he/she is registered.

NSUB_PROJECTS. This is a measure of modularity which is defined as a ‘strategy for organizing complex products and processes efficiently’. Modularity relies on system architectures that define the set of modules and their respective functions, the interfaces that allow modules to interact (compatibility) and the standards that are used to test the modules’ compliance with the design rules (Baldwin and Clark, 1997). In the case of OSS the architecture of the system typically consists of a core structure (e.g., the kernel of the Linux operating system) and a series of modules that are developed independently of one another. In large projects like Linux kernel or Apache server, a significant level of modularity is achieved through a sharp distinction between the core product architecture and more ‘external’ features that are ‘located in modules that can be selectively compiled and configured’ (Mockus et al., 2000, p. 4).

Our proxy for the project modularity is the cumulated number of subprojects with at least one opened task launched by all projects with which the individual is registered from November 1999 to January 2003. If the individual is registered to more than one project, this is the average number of sub-projects opened in all projects in which he/she is registered.

N_FILE_REL. Number of file releases (new versions of the software) from the date of registration of the project to January 2003. If the individual is registered to more than one project, this is the average number of files released of the projects in which he/she is registered.

D_LI_GPL_LGPL. Dummy equal to 1 if the main license initially adopted by the projects participated by the individual is GPL or LGPL. The dummy is equal to 0 if the main license is BSD, Public Domain, Artistic license, Apache license, MIT license, or Others. If the individual is registered with more than one project this variable indicates whether these types of license are adopted by at least one of the projects.

D_NL_ENG. Dummy equal to 1 if the official spoken language is English in at least one of the projects of the individual and it is 0 otherwise.

D_PL_ C/C++. Dummy equal to 1 if the programming language is C/C++ in at least one of the projects of the individual, and it is 0 otherwise (Java, Php, Perl, Python, Visual Basic, Unix Shell, Others).

D_OS_LINUXPOSIX. Dummy equal to 1 if the operating system is Linux or Posix in at least one of the projects of the individual, and it is 0 otherwise (Microsoft, MacOS, Independent, Others).

D_MISSING_CHAR. Dummy equal to 1 if the information on the projects' license, programming language and operating system is missing in the dataset, and equal to 0 otherwise.

3.3. Descriptive statistics

As mentioned before, the key variables of this paper are the role of individuals, their skill profile and the organization of the project in which they are active.

Table 3 summarizes the descriptive statistics of all variables at the individual level. Table 5 shows the mean and standard deviation of the main individual level variables for each of the five role profiles.

[Tables 3-4 around here]

It is worth to observe that there are substantial differences in the individual skill profile across different roles. First, the average level of skills of individuals (EXPERIENCE) increases from category 1 (No role) to 5 (PM_Oth_DEV) from 2.76 to 2.82. Roles with more complex managerial tasks (4 and 5) show also a greater variety of skills (N_SKILLS) and a more diversified skill profile (HERF). Project managers are also more likely to report their skills and to agree to publish his/her skills on the web, which suggests a higher commitment to the project and an interest to signal their profile to the OSS community.
The statistics on the other individual level variables suggests that individuals with managerial roles are more likely to show their e-mail address and a larger share of them has a company affiliation. On average they are more active in the OSS community relative to other roles, as indicated by the larger average number of messages sent to forums, and of submitted and assigned contributions.
In order to provide a deeper understanding of the skill profile of founders and project managers, Table 5 compares the skill profile of individuals who founded a project with individuals who did not found any project for the total sample and for each role category.

[Table 5 around here]

The last row of the table shows that on average founders have a more diversified skill profile and a higher skill level than non founder individuals, but this difference is very small. However, when we compare the skill profile of founders and non founders by role category we observe two interesting patterns. First, the level of skills (EXPERIENCE) of founders and non founders is the same or almost the same in each role category.

Second, the number and degree of diversification of skills is always greater for founders with respect to non founder in each role category, and this difference tends to increase in managerial roles. This suggests that founders who also assume managerial and other roles have the most diversified skill profile if compared with founders with no managerial roles or non founder with managerial roles.

Table 6 finally presents the correlation between the main regressors.

[Table 6 around here]

3.4. Regression results

We first analyze the association between individual and project level characteristics and the individuals' role by running ordered logit estimations. In the first set of regressions the dependent variable is ROLE5, a categorical and ordered variable which takes on five values corresponding to the five roles described above, which are characterized by increasing levels of managerial task complexity.

Table 7 reports the results of these estimations. Since a primary goal of this paper is the relationship between skill profile and individual's roles, we estimate three model specifications which correspond to different measures of skills. The first specification includes EXPERIENCE, in the second one we add N_SKILLS, and in the third we used HERF in place of N_SKILL. We run these three regression specifications by progressively adding to the main skill variables (columns 1-3) individual-level controls (columns 4-6), the main project-level regressors, SIZE and NSUB_PROJECTS (columns 7-9), and finally project-level controls (columns 10-12).

In line with the Lazear's theory, we find that the number of skills and the level of skill diversification (HERF) have positive and significant coefficients. Moreover the size of the coefficient of HERF is always larger than N_SKILL[2]. These results support the hypothesis that individuals who carry out multi-task jobs and take managerial responsibilities have more balanced skill sets compared with individuals who play more specialized tasks.

[Table 7 around here]

The level of skills instead is less important as a predictor of roles. The coefficient of EXPERIENCE is positive and significant only in the regressions without other individual and project level variables (columns 1-3), while it becomes insignificant in all other specifications. The descriptive statistics also show that, although the average level of skills is increasing with the managerial complexity of roles, the differences between the means across roles are very small. Project or team managers then do not necessarily need to be more skilled than specialized team members, but they must possess broader skill sets to coordinate different activities.

As far as individual-level variables are concerned, individuals who carry out managerial roles appear more open to the community, as shown by the positive and significant coefficient of D_VIEW_SKILL and of N_MESSAGES.

[2] We computed the marginal effects of the main regressors on 'managerial' roles and the results confirm a larger impact of HERF with respect to N_SKILLS. The same variables have a negative marginal effect on other roles such as 'No role' and 'Other roles'.

An interesting set of results concerns the level of activity associated with Project managers. Individuals who perform these roles submit (and are assigned) larger numbers of key tasks like bugs and patches (N_MAINCONTR_ASS and N_MAINCONTR_SUB) compared with other roles. They are also likely to submit 'minor' tasks like support requests (N_OTHERCONTR_SUB) but are unlikely to be assigned this type of tasks (N_OTHERCONTR_ASS). This is consistent with the view that in the OSS setting project managers delegate minor activities such as support requests processing to other project members. The significance of N_OTHERCONTR_ASS however decreases when all controls are included in the regressions.

SIZE and N_SUBPROJECTS are both positive and significant (columns 7-9), confirming that larger projects and projects organized in different modules require higher levels of coordination and managerial tasks. As expected, larger projects and modular design allow for a greater division of labor between specialists and more multitalented, general-purpose individuals. These results hold when controls like the type of license, programming languages, operating systems and the project spoken language are included. It is also worth noting that N_SUBPROJECT has a larger coefficient (and marginal effects) than SIZE, suggesting that managerial roles are important especially for coordinating and integrating different inputs and people working on different modules.

We also controlled for project performance by including the number of new versions of the project software and found that this is positively correlated with high managerial roles.

We also perform a second set of estimations that analyze the association between individual level characteristics and the probability that an individual founded a project. We carry out logit estimations and the dependent variable is FOUNDER. Like in the previous set of estimations, we estimate three model specifications which correspond to different measures of skills. However, in these regressions we only perform models 1 to 6 including individual variables. We do not run regressions including project-level regressors, which are not antecedent to the decision to found a project. For the same reason we also exclude from this analysis all variables that measure the level of activity of individuals in the projects of which they are members.

The results of these regressions are reported in Table 8. Our main results about the skill profile of managers are confirmed also for project founders. The coefficients of EXPERIENCE and N_SKILLS are almost identical in the two sets of regressions and the coefficient for HERF is substantially larger than the other coefficients.

Moreover, differently from the previous regressions, the coefficient for TIMEREG is positive and significant, indicating that founders registered earlier to SF.net than non founders.

[Table 8 around here]

3.5.Robustness checks

To check the robustness of our results we tried alternative classifications of roles.

At the level of individuals, we first built a 7-categories classification in which role 5 (PM and various roles) is splitted in three roles - “PM&Other”, “PM&DEV”, “PM&Other&DEV”. A 6-categories classification was also tested whereby role 5 is splitted in two roles - “PM&Other or PM&DEV”, “PM&Other&DEV”.

At the level of individual-role pairs we first employed the dataset on role-individual pairs and used ROLE4 as dependent variable. We also built a 3-category classification in which the base category is 'No Role', the second category, 'Specialist', includes all roles that perform specialized activities, and the third category, 'Generalist', includes roles that imply managerial, administrative and other support activities.

The results of estimations using these alternative classifications are very similar to the ones obtained by using ROLE5 as dependent variable[3].

A potential limitation of our estimations is about the choice of ordered logit models, which is the natural choice for categorical and ordered variables. However, if one thinks of our role categories as distinct choices independent from each other, then multinomial logit estimations should provide a better fit of the data. If this is the case, each individual, conditional upon her characteristics and the characteristics of the project, chooses one of the different role categories. To check if this model fits our dependent variables, we estimated multinomial logit models on role-individual pairs data. We estimated the full multinomial models for all alternatives and the restricted models excluding one alternative, and performed the Hausman test and the seemingly unrelated estimation (SUR) test for testing the assumption of independence of irrelevant alternatives (IIA). Both tests always rejected the IIA assumption.

These tests then support our choice of ordered logit models.[4] The roles analyzed in this paper can be naturally ordered in terms of managerial complexity and, in all likelihood, they correspond to increasingly higher positions in the team hierarchy. Moreover, over time

---

[3] Results are available upon request from the authors.

[4] We also performed ordered probit regressions and the results are qualitatively similar to those reported in the paper.

participants may gain experience in different domains and therefore become able to take more responsibility.

We did not estimate the multinomial logit model on ROLE5, because the assumption of IIA is violated by construction. It is difficult to consider independent an alternative (the fifth category) that combines other alternatives.

4. Conclusions

This paper provides novel evidence about OSS entrepreneurs. Our analysis shows that there exist marked differences between project leaders and other project members. These differences are more marked when project managers are also project founders.

Entrepreneurs carry out a larger number of different tasks, such as bug fixing and patch additions to the project software, as compared with ‘specialists’ (including ‘pure developers’). They also coordinate the job of other contributors by asking for bug or patch reports, and new features. They represent the driving force of OSS projects and have a great influence on the performance of these projects.

For these reasons it is important to understand which individual characteristics affect the emergence of OSS entrepreneurs. Our paper contributes to the literature by exploring the skill profile of founders and project leaders. We find that it is their skill diversification that distinguishes the entrepreneurs from other project members rather than the level of skills. This finding is in line with the Lazear’s theory of entrepreneurship which views entrepreneurs (and top managers) as individuals with balanced skill profiles rather than highly skilled specialists.

The association between skill profile and entrepreneurial roles is robust to control for important individual and project-level characteristics such as size and modularity.

Our results are robust to various controls and do not change significantly when other robustness checks are conducted. However, in future research we should overcome some limitations imposed by the data set. Additional individual-level information will be collected from other sources such as surveys to understand better the background of project leaders - e.g., level of education and working experience.

References

Astebro TB. 2006. Does it Pay to be a Jack of all Trades?. Available at SSRN: http://ssrn.com/abstract=925221.

Baldwin CY, Clark KB. 1997. Managing in an age of modularity. *Harvard Business Review* September-October, 84-93.

Bantel K, Jackson SE. 1989. Top management and innovations in banking: does the composition of the top team make a difference? *Strategic Management Journal* **10**(Special Issue Summer): 107-124.

Baron JN, Hannah MT. 2002. Organizational blueprints for success in high-tech start-ups. California Management Review **44**(3): 8-36.

Bhidé AV. 2000. *The origin and evolution of new business*. Oxford University Press: Oxford.

Carbonell P, Rodriguez AI. 2006. Designing teams for speedy product development: The moderating effect of technological complexity. *Journal of Business Research* 59(2): 225-232.

Cŭbranić D, Booth K. 1999. Coordinating Open Source Software Development. *IEEE Proceedings*, 8th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET ICE '99); 61–66.

Dalle JM, Jullien N. 2003. 'Libre' software: turning fads into institutions?. *Research Policy* 32(11): 1-11.

Di Bona C, Ockman S, Stone M (eds.). 1999. *Open Sources: Voices from the Open Source Revolution*. O'Reilly: Sebastopol, CA.

Galunic DC, Rodan S. 1998. Resource re-combinations in the firm: knowledge structures and the potential for Schumpeterian innovation. *Strategic Management Journal* **19**(12): 1193-1201.

Ghosh RA, David PA (2003). The nature and composition of the Linux kernel developer community: a dynamic analysis. SIEPR-Project NOSTRA Working Paper, http://dxm.org/papers/licks1/.

Hambrick DC, Cho TS, Chen M-J. 1996. The Influence of Top Management Team Heterogeneity on Firms' Competitive Moves. *Administrative Science Quarterly* **41**(4): 659-684.

Hamilton BH, Nickerson JA, Owan H. 2003. Team incentives and worker heterogeneity: an empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy* **1**(3): 465-497.

Harhoff D, Henkel J, von Hippel E. 2003. Profiting from voluntary information spillovers: how users benefit by freely revealing their innovations. *Research Policy* 32(10): 1753-1769.

Howison J, Crowston K. 2004. The Perils and Pitfalls of Mining Sourceforge. In: Proceedings of Workshop on Mining Software Repositories at the International Conference on Software Engineering ICSE Edinburgh, May 25, 2004.

Kattawattanachai P, Yoo Y. 2002. Dynamic Nature of Trust in Virtual Teams. *Journal of Strategic Information Systems* 11(3-4): 187-213.

Kayworth T, Leidner D. 2000. The Global Virtual Manager: A Prescription for Success. *European Management Journal* 18(2): 183-194.

Koch S., Schneider G. 2000. Results from Software Engineering Research into Open Source Development Projects Using Public Data. Diskussionspapiere zum Tätigkeitsfeld Informationsverarbeitung und Informationswirtschaft, Hans R. Hansen und Wolfgang H. Janko (Hrsg.), Nr. 22, Wirtschaftsuniversität Wien.

Krishnamurthy S. 2002. Cave or community? An empirical examination of 100 mature open source projects. *First Monday* **7**(6), http://firstmonday.org/issues/issue7_6/krishnamurthy/index.html.

Laursen K, Mahnke V, Vejrup-Hansen P. 2005. Do differences make a difference? The impact of human capital diversity, experience and compensation on firm performance in engineering consulting. Druid Working Paper No. 05-04; Danish Research Unit for Industrial Dynamics, Copenhagen.

Lazear EP. 2002. Entrepreneurship. *NBER Working paper* No. 9109. Cambridge: Mass.

Lazear EP. 2004. Balanced Skills and Entrepreneurship. *American Economic Review* **94**(2), 208-211.

Lerner J, Tirole J. 2002. Some simple economics of Open Source. *The Journal of Industrial Economics* 46(2): 197-234.

Lippman SA, Rumelt RO. 1982. Uncertain imitability: An analysis of interfirm differences in efficiency under competition. *Bell Journal of Economics* 13(2): 418-438.

Martins LL, Gilson LL, Maynard MT. 2004. Virtual Teams: What Do We Know and Where Do We Go From Here? *Journal of Management* 30(6): 805-835.

Mateos-Garcia J, Steinmueller WE. 2003. The Open Source Way of Working: A New Paradigm for the Division of Labour in Software Development?. SPRU - Open Source Movement Research INK Working Paper No. 1.

Milgrom P, Roberts J. 1990. The economics of modern manufacturing. *American Economic Review* 80(3): 511-28.

Mockus A., Fielding RT, Herbsleb J. 2000. A Case Study of Open Source Software Development: The Apache Server. Proceedings of the Twenty-Second International Conference on Software Engineering: 263–272.

Raymond ES. 2001. *The Cathedral and the Bazaar. Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly: Sebastopol, CA.

Raymond ES. 1999. Linux and open-source success. IEEE Software **16**(1): 85-89.

Schumpeter J. 1942. *Capitalism, Socialism and Democracy*. Unwin University Books: London.

Sutton RI, Hargadon A. 1997. Brainstorming groups in context: effectiveness in a product design firm. *Administrative Science Quarterly* **42**(4): 685-718.

von Hippel E. 2001. Learning from open source software. *Sloan Management Review* **42**(4): 82-86.

Wagner J. 2003. Testing Lazear's "Jack-for-All-Trades View of Entrepreneurship with German Micro Data, *Applied Economic Letters* **10**(11): 687-689.

Wagner J. 2006. Are Nascent Entrepreneurs Jack-of-All Trades? A Test of Lazear's Theory of Entrepreneurship with German Data. *Applied Economics* 38(20): 2415-2419.

Table 1. Distribution of pairs by role category (ROLE4)

	Role	N	%
1	No role	54,879	51.37
2	Other roles (Other)	10,647	9.97
3	Developer (DEV)	30,009	28.09
4	Project Manager (PM)	11,288	10.57
	Total pairs	**106,823**	**100**

Table 2. Distribution of individuals by role category (ROLE5)

	Role	N	%	% founders *
1	No role	34,351	44.59	82.37%
2	Other roles	7,885	10.24	30.36%
3	Developer (DEV)	24,581	31.91	23.55%
4	Project Manager (PM)	8,081	10.49	70.21%
5	PM &/or Other &/or DEV	2,141	2.78	81.98%
	Total individuals	**77,039**	**100**	**55.48%**

* the number of individuals for which information about the founder state is available is 66944.

Table 3. Descriptive statistics – key regressors and controls

Variable	Mean	Std. Dev.	Min	Max
*EXPERIENCE**	2.770	0.734	1	5
*1-HERF**	0.724	0.216	0	0.97
*N_SKILLS**	6.605	3.607	1	29
D_MISSING_SKILLS	0.694	0.461	0	1
D_VIEW_SKILLS	0.134	0.341	0	1
TIME_REG	18.556	10.530	1	40
N_PROJECTS	1.387	0.995	1	52
D_MAIL	0.197	0.398	0	1
D_MAIL_COM	0.082	0.274	0	1
N_MESSAGES	2.875	23.924	0	1758
N_MAINCONTR_SUB	1.430	8.852	0	1034
N_MAINCONTR_ASS	1.473	15.262	0	976
N_OTHERCONTR_SUB	0.490	2.636	0	448
N_OTHERCONTR_ASS	0.621	75.474	0	19052
SIZE	5.937	9.300	1	102
NSUB_PROJECTS	0.692	1.730	0	45
N_FILE_REL	4.673	11.773	0	444
D_NL for English	0.813	0.390	0	1
D_NL for German	0.063	0.244	0	1
D_NL for French	0.035	0.184	0	1
D_NL for Italian	0.013	0.113	0	1
D_NL for Spanish	0.018	0.132	0	1
D_NL for Portuguese	0.010	0.100	0	1
D_NL for Russian	0.010	0.099	0	1
D_NL for Polish	0.005	0.074	0	1
D_NL for Chinese	0.005	0.072	0	1
D_NL for Japanese	0.005	0.073	0	1
D_NL for Swedish	0.006	0.076	0	1
D_NL for Dutch	0.007	0.085	0	1
D_NL for Other	0.009	0.094	0	1
D_LI_GPL_LGPL	0.809	0.393	0	1
D_OS_LINUXPOSIX	0.636	0.481	0	1
D_PL_ C/C++	0.610	0.488	0	1
D_NL_ENG	0.818	0.386	0	1
D_MISSING_CHAR	0.239	0.405	0	1

Note: Number of observations =77,039.
* For the variables using information about skills the number of observations is 23,560.

Table 4. Profile of individuals by role

Variable	(1) No role	(2) Other	(3) DEV	(4) PM	(5) PM_ Other_DEV	Total
*EXPERIENCE**	2.756	2.754	2.764	2.823	2.815	2.770
	(0.736)	(0.759)	(0.737)	(0.713)	(0.681)	(0.734)
*1-HERF**	0.710	0.713	0.730	0.740	0.770	0.724
	(0.226)	(0.221)	(0.212)	(0.200)	(0.175)	(0.216)
*N_SKILLS**	6.351	6.579	6.664	6.879	7.540	6.605
	(3.555)	(3.682)	(3.582)	(3.617)	(3.746)	(3.607)
D_MISSING_SKILLS	0.728	0.677	0.704	0.607	0.428	0.694
	(0.445)	(0.468)	(0.456)	(0.488)	(0.495)	(0.461)
D_VIEW_SKILLS	0.106	0.154	0.135	0.184	0.318	0.134
	(0.308)	(0.361)	(0.341)	(0.387)	(0.466)	(0.341)
TIME_REG	20.420	16.362	16.402	18.040	23.422	18.556
	(10.975)	(9.867)	(9.743)	(9.917)	(9.299)	(10.530)
N_PROJECTS	1.208	1.333	1.459	1.407	3.531	1.387
	(0.618)	(0.988)	(1.036)	(0.921)	(2.286)	(0.995)
D_MAIL	0.144	0.194	0.211	0.288	0.554	0.197
	(0.351)	(0.395)	(0.408)	(0.453)	(0.497)	(0.398)
D_MAIL_COM	0.059	0.085	0.088	0.122	0.223	0.082
	(0.235)	(0.279)	(0.283)	(0.327)	(0.416)	(0.274)
N_MESSAGES	1.545	3.166	2.893	5.642	12.505	2.875
	(11.678)	(20.333)	(25.224)	(37.809)	(63.531)	(23.924)
N_MAINCONTR_SUB	0.702	1.559	1.684	2.127	7.089	1.430
	(3.998)	(10.096)	(9.240)	(9.733)	(28.407)	(8.852)
N_MAINCONTR_ASS	0.473	1.149	1.682	2.982	10.631	1.473
	(7.114)	(10.449)	(15.663)	(21.842)	(49.413)	(15.262)
N_OTHERCONTR_SUB	0.338	0.545	0.378	0.883	2.517	0.490
	(1.263)	(5.612)	(1.900)	(2.413)	(6.304)	(2.636)
N_OTHERCONTR_ASS	0.352	0.512	0.216	0.478	10.544	0.621
	(44.365)	(24.202)	(8.936)	(11.581)	(412.051)	(75.474)

Note: Number of observations = 77,039. Standard errors in parenthesis.
* For the variables using information about skills the number of observations is 23,560.

Table 5. Skill profile of founders by role

	N		EXPERIENCE		N_SKILLS		1-HERF	
	Founder	No_Founder	Founder	No_Founder	Founder	No_Founder	Founder	No_Founder
No role	6769	979	2.75	2.73	6.40	6.10	0.71	0.70
			(0.72)	(0.81)	(3.57)	(3.51)	(0.22)	(0.23)
Other roles	1206	1113	2.73	2.74	7.09	5.91	0.74	0.68
			(0.73)	(0.80)	(3.84)	(3.41)	(0.21)	(0.24)
Developer	2520	3961	2.74	2.74	6.91	6.43	0.74	0.72
			(0.69)	(0.76)	(3.70)	(3.46)	(0.20)	(0.22)
Project Manager	2556	418	2.81	2.82	6.96	6.48	0.74	0.71
			(0.71)	(0.72)	(3.61)	(3.68)	(0.20)	(0.22)
PM_Other_DEV	1009	133	2.81	2.81	7.67	6.67	0.78	0.72
			(0.66)	(0.77)	(3.74)	(3.82)	(0.17)	(0.21)
Total	14060	6604	2.76	2.74	6.74	6.30	0.73	0.71
			(0.71)	(0.77)	(3.66)	(3.49)	(0.21)	(0.23)

Note: standard errors in parenthesis. For this Table we used a sample of 20,664 individuals for which we have information on skills and on founder status (whether they have founded a project or not).

Table 6. Correlations between main regressors

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]
[1] EXPERIENCE	1														
[2] 1-HERF	-0.69*	1													
[3] N_SKILLS	0.80*	-0.33*	1												
[4] D_VIEW_SKILLS	0.53*	-0.30*	0.56*	1											
[5] TIME_REG	0.14*	-0.07*	0.13*	0.08*	1										
[6] D_MAIL	0.18*	-0.09*	0.17*	0.14*	0.20*	1									
[7] D_MAIL_COM	0.10*	-0.05*	0.10*	0.09*	0.11*	0.60*	1								
[8] N_MESSAGES	0.06*	-0.03*	0.06*	0.06*	0.06*	0.15*	0.11*	1							
[9] N_MAINCONTR_SUB	0.07*	-0.03*	0.07*	0.06*	0.10*	0.28*	0.18*	0.15*	1						
[10] N_MAINCONTR_ASS	0.06*	-0.02*	0.04*	0.03*	0.08*	0.19*	0.13*	0.24*	0.39*	1					
[11] N_OTHERCONTR_SUB	0.09*	-0.04*	0.08*	0.06*	0.11*	0.24*	0.16*	0.14*	0.23*	0.21*	1				
[12] N_OTHERCONTR_ASS	0.01*	-0.01*	0.01*	0.01*	0.01	0.02*	0.02*	0.04*	0.04*	0.11*	0.10*	1			
[13] SIZE	-0.03*	0.03*	-0.04*	-0.01*	-0.03*	0.05*	0.04*	0.03*	0.04*	0.04*	0.02*	0.00	1		
[14] NSUB_PROJECTS	0.03*	-0.01*	0.04*	0.04*	0.01*	0.07*	0.05*	0.05*	0.06*	0.04*	0.02*	0.00	0.35*	1	
[15] N_FILE_REL	0.03*	-0.01*	0.01*	0.02*	0.05*	0.19*	0.12*	0.09*	0.09*	0.09*	0.08*	0.00	0.37*	0.22*	1

*p <0.05.

Table 7. Results of the ordered logit estimations. Dependent variable: ROLE5

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
EXPERIENCE	0.070***	0.068***	0.065***	0.024	0.024	0.022	0.015	0.015	0.012	0.027	0.026	0.024
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
D_MISSING_SKILLS	-0.164***	0.083	-0.338***	-0.152***	0.016	-0.306***	-0.206***	-0.040	-0.361***	-0.062	0.109*	-0.214***
	(0.048)	(0.054)	(0.051)	(0.050)	(0.055)	(0.053)	(0.050)	(0.055)	(0.053)	(0.051)	(0.055)	(0.054)
N_SKILLS		0.038***			0.028***			0.028***			0.029***	
		(0.004)			(0.004)			(0.004)			(0.004)	
1-HERF			0.584***			0.466***			0.470***			0.462***
			(0.057)			(0.059)			(0.059)			(0.060)
D_VIEW_SKILLS				0.257***	0.216***	0.221***	0.241***	0.201***	0.205***	0.196***	0.154***	0.160***
				(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.026)	(0.026)
TIME_REG				-0.031***	-0.031***	-0.031***	-0.030***	-0.030***	-0.030***	-0.024***	-0.024***	-0.024***
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
D_MAIL				0.656***	0.651***	0.652***	0.619***	0.614***	0.615***	0.499***	0.493***	0.495***
				(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.027)	(0.027)	(0.027)
D_MAIL_COM				-0.024	-0.023	-0.023	-0.037	-0.036	-0.036	-0.021	-0.020	-0.021
				(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.034)	(0.034)	(0.034)
N_MESSAGES				0.003***	0.003***	0.003***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
N_MAINCONTR_SUB				0.007***	0.007***	0.007***	0.006***	0.005***	0.006***	0.005***	0.005***	0.005***
				(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
N_MAINCONTR_ASS				0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N_OTHERCONTR_SUB				0.064***	0.064***	0.064***	0.070***	0.069***	0.069***	0.066***	0.066***	0.066***
				(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
N_OTHERCONTR_ASS				-0.000*	-0.000*	-0.000*	-0.000	-0.000*	-0.000*	-0.000	-0.000	-0.000
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
NSUB_PROJECTS							0.078***	0.078***	0.078***	0.056***	0.056***	0.056***
							(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
SIZE							0.023***	0.024***	0.024***	0.020***	0.020***	0.020***
							(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N_FILE_REL										0.004***	0.004***	0.004***
										(0.001)	(0.001)	(0.001)
Observations	77039	77039	77039	77039	77039	77039	77039	77039	77039	77039	77039	77039
Log Pseudolikelihood	-99379.83	-99314.78	-99325.41	-97122.59	-97089.07	-97089.94	-96163.69	-96131.02	-96130.73	-93980.63	-93946.51	-93949.66
Wald chi2	588.56	688.67	680.5	4175.35	4217.37	4223.24	5598.86	5642.31	5651.67	10148.81	10189.7	10188.32

* $p<0.10$, **$p<0.05$, ***$p<0.01$. Robust standard errors in parenthesis.

Regressions from column (4) to column (12) include dummies for the natural language of the individual (*D_NL*).

Regressions from column (10) to column (12) also include the following dummies: *D_NL_ENG, D_OS_LINUXPOSIX, D_PL_ C/C++, D_LI_GPL_LGPL, D_MISSING_CHAR.*

Table 8. Results of the logit estimations. Dependent variable: FOUNDER

Variable	(1)	(2)	(3)	(4)	(5)	(6)
EXPERIENCE	0.04**	0.04*	0.04*	-0.01	-0.01	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
D_MISSING_SKILLS	-0.65***	-0.43***	-0.79***	-0.59***	-0.47***	-0.68***
	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.07)
N_SKILLS		0.03***			0.02***	
		(0.00)			(0.00)	
1-HERF			0.46***			0.26***
			(0.07)			(0.07)
D_VIEW_SKILLS				0.05	0.02	0.03
				(0.03)	(0.03)	(0.03)
TIME_REG				0.04***	0.04***	0.04***
				(0.00)	(0.00)	(0.00)
D_MAIL				0.42***	0.41***	0.42***
				(0.03)	(0.03)	(0.03)
D_MAIL_COM				0.08*	0.08**	0.08**
				(0.04)	(0.04)	(0.04)
_cons	0.64***	0.42***	0.32***	-0.19***	-0.31***	-0.35***
	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)	(0.08)
N	66944	66944	66944	66944	66944	66944
Ll	-45023.9	-44989.9	-45000.9	-43362.8	-43351	-43356.3
chi2	1875.51	1924.72	1915.09	5163.99	5177.04	5173.88

* p<0.10, **p<0.05, ***p<0.01. Robust standard errors in parenthesis.
Regressions from column (4) to column (6) include dummies for the natural language of the individual (*D_NL*).

Appendix

Table A.1. Roles reported in SF.Net

Role	Share
Developer	28.09%
Project Manager	10.57%
All-Hands Person	3.77%
Web Designer	1.02%
Tester	0.85%
Graphic/Other Designer	0.77%
Advisor/Mentor/Consultant	0.58%
Doc Writer	0.57%
Doc Translator	0.46%
Unix Admin	0.42%
Analysis / Design	0.34%
Packager (.rpm, .deb etc)	0.29%
Editorial/Content Writer	0.21%
Porter (Cross Platform Devel.)	0.19%
Content Management	0.17%
Support Manager	0.16%
Distributor/Promoter	0.09%
Requirements Engineering	0.08%
No role	51.37%
Total	100.00%

Number of observations = 106,823.