

What Makes Geeks Tick?

A Study of Stack Overflow Careers

Lei Xu^{*1}, Tingting Nian^{†2}, and Luis Cabral^{‡3}

¹McGill University

²NYU Stern School of Business

³NYU Stern School of Business

December 15, 2014

**PRELIMINARY AND INCOMPLETE
PLEASE DO NOT CIRCULATE!**

Abstract

Online communities such as Stack Overflow provide an ideal setting for studying incentives on the voluntary contribution to public goods. One particularly intriguing question is, to what extent such contribution is driven by career concerns. We estimate the magnitude of the impact of career concerns on the contribution to Stack Overflow using a set of users who experience a job change, which also brings a change in career concerns. We use difference-in-differences method to tease out potential confounding factors such as changes in time availability which can also cause changes in contribution levels. We estimate that after changing to a new job, online contribution level decreases by 20.3%, in which 14.5 percentage points drop is due to career concerns.

Keywords: career concerns, signaling, labor supply, online contribution, intrinsic motivation, extrinsic motivation

JEL Classification Numbers: L14, H41, J22, J24, M51, M53, D83

^{*}Visiting Scholar, Business Economics, Rotman School of Management, University of Toronto; Ph.D. Candidate, McGill University. Contact: lxu@stern.nyu.edu

[†]Ph.D. Candidate, Department of Information, Operations & Management Sciences, NYU Stern School of Business. Contact: tnian@stern.nyu.edu

[‡]Paganelli-Bull Professor of Economics and International Business, Stern School of Business, New York University. Contact: luis.cabral@nyu.edu

1 Introduction

The Internet has revolutionized the world in more than one way. Of particular interest is the phenomenon of the private contribution to collective projects such as Wikipedia, bulletin boards, or open source software development. As Lerner and Tirole (2002) put it, to an economist the behavior of individual contributors seems a bit startling: is there a case of altruism, or are there ulterior motives behind private contributions to a public good?

Our paper addresses the question from an empirical approach. We use data from Stack Overflow, one of the most important online question boards for programming related matters. We consider a hypothesis put forward by Lerner and Tirole (2002), namely that contributions are motivated by career concerns.

There is a site associated with Stack Overflow, called Stack Overflow Careers, which lists contributors' CVs. Access to the site by employees is by invitation only.¹ Stack Overflow Careers can be accessed by potential employers upon payment of a fee. By paying such fee, they gain access to all user profiles as well as their contribution activity on Stack Overflow. Stack Overflow Careers makes the hiring process more efficient not only through the ability to search for candidates with the exact skills needed for a job, but also through the better screening process when selecting candidates for next-round interview. A tantalizing possibility — the hypothesis we propose to test — is that contributing to Stack Overflow is a way of signaling one's ability with a view at being able to find a better job.

With contributors' profiles on Stack Overflow Careers and their historical activities on Stack Overflow, we are able to construct complete histories of each individual's online trajectory: contribution to Stack Overflow as well as individual characteristics and employment history. Based on this data, we can directly test the career concerns hypothesis by identifying shifts in behavior following career-relevant shifts, namely a change of employment.

The goal of this paper is to empirically examine the causal relationship between extrinsic motivations and contribution to the public good. By investigating this relationship, our research contributes to the broad literature on the motivations for voluntary work, as well as research on career concerns. Previous studies have widely documented a number of motivations for voluntary contribution. However, as far as we are aware, there have been no studies that could empirically separate the impact of career concerns on the contribution to the public good. We believe the question is of particular interest given the prevalence of collective collaboration on user-generated content in various online communities such as Wikipedia, Stack Overflow, and YouTube.

¹The invitation is — at least to some extent — a function of a contributor's online activity and expertise in a certain area.

2 Literature Review

There is a broad theoretical literature on why people contribute to the public good, especially Open Source Software (OSS). In essence, helping others to solve programming problems on Stack Overflow is very similar to working on an OSS project. Both of them are done in a voluntary fashion (i.e. without monetary compensation), and the work is shared freely online to the public.

Lerner and Tirole (2001) analyze the incentives for people to participate in OSS. Apart from altruistic reasons, they argue that the motivations provided by career concerns can be equally strong; e.g. frequent open source contributors, such as Sun and Netscape, had more access to venture capital. They also invited scholars for more empirical and theoretical research on the incentives in OSS, especially career concerns and ego gratification, which are driven by the fact that their contribution is visible to special audiences.

von Krogh et al. (2012) draws a thorough review of the literature discussing the motivation to contribute to OSS projects, and summarizes the literature into three groups: Intrinsic (e.g. Altruism, Ideology, Fun, Kinship), Internalized Extrinsic (e.g. Reputation, Learning, Reciprocity, Own-Use), and Extrinsic (e.g. Career Concerns, Pay). Our study attempts to empirically separate the impact of extrinsic motivations from that of intrinsic motivations on the provision to public goods.

Among the few empirical papers on the incentive to contribute to public goods, Bitzer and Geishecker (2010) analyze which individual characteristics of IT employees are associated with the willingness to participate voluntarily in OSS projects using survey data, and find that formal educational attainment is not positively associated with the probability or level of OSS contributions. Moreover, university dropouts show a particularly high probability of working on such projects. The authors interpret this as evidence of career-oriented motives in OSS contributors.

Roberts, Hann and Slaughter (2006) use a longitudinal field study which combined archival data from OSS project records and a targeted survey to examine the disparate impacts that different motivations (intrinsic, internalized extrinsic, and extrinsic) have on the level of participation. While some research has focused on motivations which are antecedents of performance and behavior, others study the consequences of participation in OSS projects.

Drawing upon theories such as job signaling, some researchers have posited there are economic returns to the participation in OSS projects since a user's participation might act as a credible "signal" of her quality. Among the very few empirical studies, Hann, Roberts and Slaughter (2013) find empirical evidence of financial returns associated with programmers' participation in OSS projects by following a cohort of participants overtime.

Zhang and Zhu (2011) examine the causal relationship between group size and incentives to contribute using data from Chinese Wikipedia. Their identification

comes from the exogenous reduction in group size at Chinese Wikipedia as a result of the blocking of Chinese Wikipedia in mainland China in October 2005. They find a significant reduction of contribution levels by nonblocked contributors as a result of the blocking. They attribute the finding to a theory that contributors benefit from more readership, which, in von Krogh et al. (2012)’s categorization, is part of “Intrinsic” or “Internalized Extrinsic” motivation.

Lerner and Tirole (2002) summarize early career concerns literature from Holmström (1999), saying that the signaling incentive is stronger when (i) performance is more visible, (ii) effort has a stronger impact on performance and (iii) performance is more informative about talent. They also highlight the two distinct, but hard-to-distinguish incentives: career concerns and ego gratification. They emphasize the distinction of these two incentives “provide lenses through which the structure of open source projects, the role of contributors, and the movement’s ongoing evolution can be viewed.” Our paper will effectively separate these two effects and focus on the effect of career concerns on the incentive to contribute.

Blatter and Niedermayer (2008) provide a theoretical model where developers choose to work on an open source project, or on a closed source project. They show that under certain conditions, a talented employee may initially prefer a lower paying job as an open source developer to commercial closed source projects, because a publicly available signal gives him a better bargaining position when negotiating wages with her employer. Their paper provides a plausible explanation, which is confirmed by the evidence from Stack Overflow.

Spiegel (2009) examines the incentive of unpaid contributors to contribute to OSS projects in order to signal their talents using a theoretical model. He then solves for the equilibrium effort level and success/failure of the product. His focus actually highlights the difference between OSS and Stack Overflow. OSS might succeed or fail, but Stack Overflow users always benefit from more answers.

However, as far as we are aware, there have been no studies that empirically tested the effect of career concerns on the contribution to the public good. At the same time, there are a few papers that empirically test the effect of career concerns on behavior in other settings.

Chevalier and Ellison (1999) examine career concerns in mutual fund managers by looking at how the likelihood of a manager being “terminated” is affected by the manager’s actions and past performance. They find that “termination” is more performance-sensitive for younger managers, and they have an incentive to avoid unsystematic risk due to the termination-performance relationship.

Our paper analyzes the impact of *extrinsic* incentives (career concerns) on contribution levels. Complementary to our question, Kolstad (2013) measures the effect of *intrinsic* incentives in the field of medicine, specifically regarding surgeons. He isolates the relative role of extrinsic and intrinsic incentives from

the introduction of quality “report cards” for cardiac surgery in Pennsylvania, and finds evidence of the presence of non-pecuniary incentives resulting from quality reporting.

3 Theoretical Model

We propose a simple dynamic model of user contributions. Consider a discrete time, infinite period model with discount factor δ . Each user has a measure of time T to distribute between three types of tasks: $\tau = w, e, a$, where τ is a generic descriptor of task type. w denotes work, e edits on Stack Overflow and a answers on Stack Overflow. Although there are mainly three different activities on Stack Overflow - asking questions, answering questions and editing others’ posts, the voluntary activities only involve the last two activities since a user posts questions mostly because she wants to learn the answer. The main difference between the edit task and the answer task is that the latter generates reputation points for the user.

There are two possible states of a user’s current job status, $s \in \{0, 1\}$, where 0 could refer to the old job/bad job, while 1 indicates the new job/good job. Here, we make a simplifying assumption that $s = 1$ is an absorbing state, which we will relax in a more general Markov model with multiple states later on. The transition probability from state 0 to state 1 is a function of a user’s reputation points at that time period. We denote it by $p(r_t)$. The user’s reputation points is in turn a function of the user’s activities that generate reputation points.

The total measure of type τ tasks available is given by m_τ . The utility of a given w task, u_w , is distributed according to cdf F_s ; the utility of a given edit and answer, u_e and u_a , is distributed according to G_s . We allow for F and G to be state dependent but require the distributions of e and a tasks to vary in tandem. More over, we require F_1 (resp. G_1) to dominate F_0 (resp. G_0) in the sense of first-order stochastic dominance (Milgrom (1981)). For simplicity, we assume that $F_s(0) = 0$ and $G_s(0) = 0$, that is, all activities have strictly positive value.

The timing of this model is as follows: first, nature decides which state s a user is in; next, nature decides the value of each type of tasks that a user could exert effort in. We also make the following assumptions:

- 1) The vote generating activity, a_t , improves a user’s reputation;
- 2) A user’s reputation is equal to the vote generating activity a_t in the previous period;
- 3) Higher reputation increases the probability of state switch: $p(r_t)$;
- 4) $F_1(w)$ first-order stochastically dominates $F_0(w)$;
- 5) The relative intrinsic value of edit and answer tasks does not depend on state

Assumption 2 is a simplifying one that a user’s reputation only depends on her activities in the previous period. Assumption 4 points out that the new job is more attractive than the old job that it has a higher value. Assumption 5 depicts the idea that the relative intrinsic value of edit and answer tasks derived from altruism and ego gratification excluding career concerns should not change from state 0 to state 1, while the absolute value may be state dependent. This assumption comprises our core identification strategy. With these assumptions, we have the following proposition.

Proposition 1. *Suppose that $F_0(w) > F_1(w)$, then $a_t|_{s=1} < a_t|_{s=0}$; Moreover, $\frac{a_t}{e_t}|_{s=1} < \frac{a_t}{e_t}|_{s=0}$, iff $p'(\cdot) > 0$*

The proof of Proposition 1 can be found in the appendix. Proposition 1 states that if $F_1(w)$ first-order stochastically dominates $F_0(w)$, as a user switches from state 0 to state 1, his activity in answering tasks should decline. However, this may be due to a lot of reasons, such as time availability that the new job simply demands more time from the user. Proposition 1 further states that the ratio of answer to edit activity should also decrease if and only if the state transition probability is increasing in the user’s reputation, a.k.a. $p'(\cdot) > 0$. This is the career element that we are interested in.

4 Data

We obtain our dataset from Stack Overflow and Stack Overflow Careers. Stack Overflow is a platform where users can ask and answer questions. Stack Overflow is the largest programming site where programmers ask and answer programming-related questions. It has a system of voting, badges, Wikipedia-style editing and reputation that ensures high-quality, peer-reviewed answers. Stack Overflow is well known by both contributors and programming-related companies.

Stack Overflow Careers is a related employment website that posts job listings as well as resumes. For contributors, building a career profile on the website is free of charge but by invitation only, and the invitation is based on the contributors’ quality of contribution and activity. On the resume, contributors can link it to their Stack Overflow profile, through which employers can learn more about their expertise. For employers, access to job candidate information requires a subscription. With the subscription, employers can post their jobs and search for job candidates. Besides the basic information of education and work experience, the link to Stack Overflow profiles gives employers detailed information on job candidates’ expertise.

Stack Overflow Careers helps employers by reducing the search cost towards hiring new employees through the following two channels: (i) Applicants are pre-

Nicholas Larsen

Norcross, GA, United States
stackoverflow.com
@fody



Last seen 2 days ago

Top 10% for asp.net asp.net-mvc asp.net-mvc-3

Top 20% for c# javascript jquery css algorithm more

Currently **Software Developer** at **Stack Overflow**.

Technologies

Likes: design-patterns algorithm-design artificial-intelligence prototyping database-design

Experience [show all](#)

Software Developer, *Stack Overflow* January 2011 - Current

asp.net-mvc c# sql-server performance redis dapper mini-profiler internationalization elasticsearch

Database Programmer, *Credit Union Service Corporation* March 2009 - December 2010

asp.net c# sql-server oracle crystal-reports route-map visual-basic c++ jquery

Education [show all](#)

Computer Science - Databases and Knowledge Systems, *Georgia State University* 2001 - 2008

java c++ ruby-on-rails databases game-theory algorithms modeling electronics embedded-systems

Stack Exchange [show all](#)

Last seen 2 days ago

Accounts

Stack Overflow **10084** reputation points

Meta Stack Exchange **7914**

Top Answers

[MVC and NOSQL: Saving View Models directly to MongoDB?](#) 6 votes
asp.net-mvc mongodb separation-of-concerns

[LINQ union with optional null second parameter](#) ✓ 6 votes
c# linq union

Figure 1: Sample Profile on Stack Overflow Careers

selected by Stack Overflow Careers. In other words, since only high-level contributors are invited, employers are presented with an elite applicant pool. (ii) Because employers can check applicants' contributions on Stack Overflow, employers are able to gain a better understanding of applicants' ability.

4.1 Data Selection

To estimate the impact of career concerns on contribution activity, we focus on users with the following criteria:

- U.S. Users: Most jobs on Stack Overflow Careers are located in the United States, and it's more common for US employers to ask for job applicants' Stack Overflow profiles.
- Job Changers: The change of career concerns comes from a job change. We select users who experienced a job change from January 2011 to July 2014.
- Job Switchers: Employment status (employed vs. unemployed) introduces unnecessary noises. We select users who are *switching* from one job to another, with a gap less or equal than one month.
- Active Users: For many users, we do not observe any activity on Stack Overflow during periods of the job change. For more accurate estimation, we focus on active users, defined as having at least one answer and at least one edit activity within four-month before and after the month of job change.

4.2 Contribution Activity on Stack Overflow

Answers Any registered user can provide answers to others' questions. Good answers are voted up by other users and rise to the top. Users earn reputation score when others vote up their answers. In order to take into consideration both quality and quantity of answers, we measure users' answers activity by counting the total votes gained from those answers in a given month.

Edits Apart from casting votes, users can also make or suggest minor changes to questions and answers. Those edits help to make the questions and answers more readable and understandable to future viewers. Same as writing answers, this type of activity can also be time consuming. However, it does not reward with any reputation points to contributors, that is to say, people contribute purely out of altruism. We sum up the total number of edits made by a contributor in a given month, which is denoted by *Edits*.

5 Identification Strategy

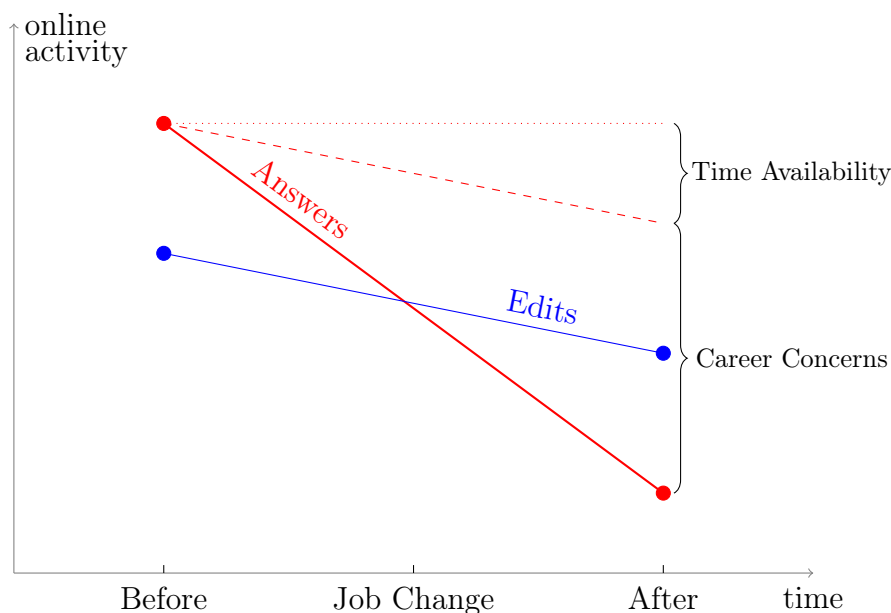


Figure 2: Identification: Difference-in-Differences

Online activity is part of the signals sent by job seekers to employers. So job seekers have incentive to maintain an outstanding online profile through giving high-quality answers. After getting a new job, career incentives of online activity has disappeared, which causes a reduction in online contribution level.²

However, part of the reduction in online activity might come from a change in time availability.³ To take this into account, we conduct difference-in-differences regressions using *Edits* as a control group which proxies for time availability. Answering questions generates votes, and it is affected by extrinsic motivation such as career concerns, whereas editing others' posts is caused by intrinsic motivation and can reflect a change in time availability.

The parallel trend assumption required for Difference-in-Differences is essentially the same as one of the assumptions we listed for the theoretical model, which is that the relative intrinsic value of edit and answer tasks does not depend on state, that is to say, if there were no career concerns, then the ratio of the number of edits and answers does not change before and after starting a new job.

²It might not totally disappear, because job seekers still have incentives to build reputation in order for job hunting in the future, in which case our results could be underestimated.

³A new job often requires time on training and familiarity with the new environment.

The Difference-in-Differences identification strategy is illustrated in Figure 2. After starting a new, the reduction of answer activity comes from two sources: 1) Career concerns 2) Time availability. But the reduction of edit activity has only one source which is the time availability.

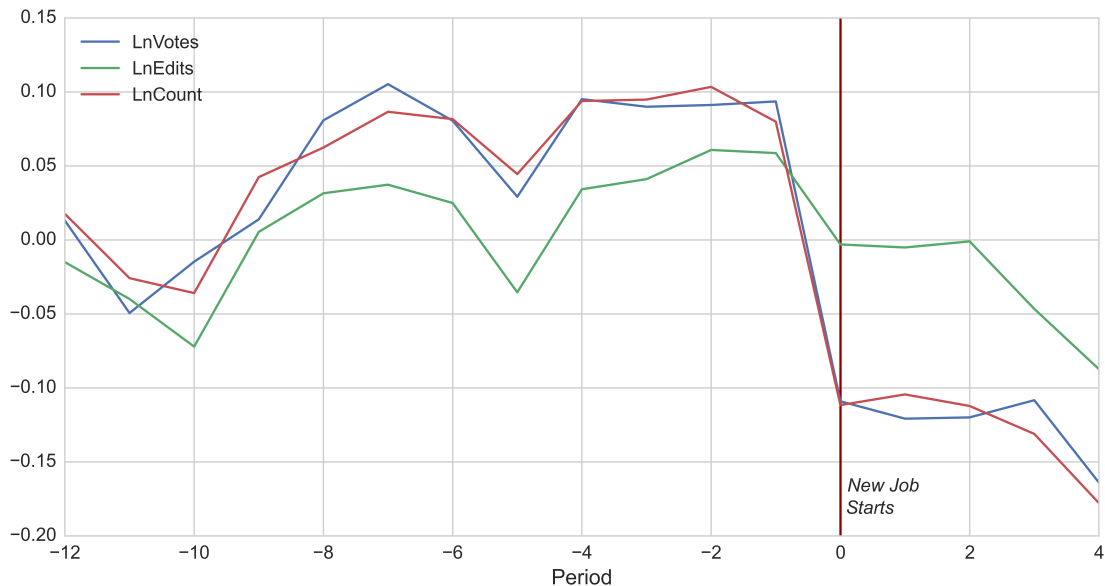


Figure 3: Average Monthly Activity by Period

Figure 3 gives a preliminary plotting of logarithm of average monthly activity. Period 0 is the month when contributors start a new job. Figure 3 shows monthly activity 12 months before and 4 months after the job change. The answer activity is represented by two different lines: Count and Votes. Count is the number of answers in a given month by a contributor. Votes measures the total votes gained those answers within 30 days after the date of each answer.

In Figure 3, the fluctuation of *Votes* and *Count* are almost identical, which means that there is no significant change in average votes per answer over time. Both answer and edit activities experience a drop when starting a new job at Period 0, though the drop in answer activity is much more significant compared to the drop in edit activity.

0	month when new job starts
a_0	log # answers in months -4, -3, -2 (monthly average)
a_1	ditto for months 1, 2, 3
e_0, e_1	ditto for edits
x_s	generic activity: $x \in \{a, e\}$
s	1 if new job, 0 otherwise
i	user index

Note: each user corresponds to one value each of a_0, a_1, e_0, e_1

Table 1: Notation for Regression 1 and 2

6 Regression Analysis⁴

Through Stack Overflow Careers profiles, employers can select the right candidates by tracing back to the contribution activities on Stack Overflow. Job candidates with high reputation and highly-voted answers send a positive signal to potential employers.

Our hypothesis is that career concerns motivate a job candidate to improve the signal of her ability to employers through contribution activities on Stack Overflow. If this is the case, then after acquiring a new job, career concerns diminishes and the contribution level drops.

To evaluate this hypothesis, we collect a panel of contributors' dates of changing jobs and their associated Stack Overflow IDs. With their IDs, we collect their contribution activity during four-month before and four-month after the month of job change. Then we use the following regression to estimate the impact of career incentive on online voluntary contribution:

$$a_{is} = \beta_1 + \beta_2 s + \epsilon_{is} \quad (1)$$

$$x_{is} = \beta_1 + \beta_2 s + \beta_3 \mathbf{1}(x = a) + \beta_4 s \mathbf{1}(x = a) + \epsilon_{is} \quad (2)$$

where

$$i \in \{1, \dots, 780\} \quad \text{as of December 15, 2014}$$

$$s \in \{0, 1\}$$

$$x \in \{a, e\}$$

Regression function (1) measures the change of answer activity after a contributor experiences a job change. Regression (2) uses a Difference-in-Differences specification explained in previous section. The dependent variable, x_{is} , includes

⁴For a more complete regression analysis and graphs, please refer to the Online Appendix at <http://leixu.org/>

	a_{is}	a_{is}	x_{is}	x_{is}
NewJob($s=1$)	-0.203*** (-4.18)	-0.203*** (-5.68)	-0.0587 (-1.31)	-0.0587** (-2.12)
Answer($x=a$)			0.483*** (10.75)	0.483*** (16.04)
NewJob \times Answer			-0.145** (-2.28)	-0.145*** (-5.49)
Constant	1.054*** (30.60)	1.054*** (58.82)	0.571*** (17.96)	0.571*** (29.76)
<i>Model</i>	OLS	FE	OLS	FE
<i>N</i>	1560	1560	3120	3120
<i>R</i> ²	0.0111	0.0397	0.0572	0.127

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Difference-in-Differences Estimation: Quantity of Answer and Edit Activities

two two types of activities: *Answers* ($x = a$) and *Edits* ($x = e$). Both types are measured by the monthly average number of activities incurred. s indicates whether the current state is 0 (old job) or 1 (new job). In regression (2), β_4 captures the differences in the changes of answer and edit activities before and after a job change. Table (2) summarizes the results from both regression (1) and (2).

Results from the first two columns says that after a job change, answer activity drops by 19.8%. However, the Difference-in-Differences regressions in the last two columns show that edit activity drop by 5.8%, and answer activity drops by an additional 14%. If the assumptions from the theoretical model is true, then we can conclude that within the total reduction of 19.8% in answer activity, 5.8 percentage point is due to a change in time availability and the other 14 percentage point drop is due to a change career concerns.

In the theory of career concerns, the future prospects for employment depend on the agent's current performance. Therefore, an agent has incentive to improve her signal to employers. All the questions and answers on Stack Overflow are peer-reviewed through vote-casting. Employers evaluate job candidates not only through the breath of their expertise, but also through the depth of their knowledge, which can be reflected by the votes casted by other users. Therefore, the signal of expertise should be reflected by both the quantity and quality of answer activity on Stack Overflow. A user probably gains more reputation points from one well-written answer than five poorly-written answers. In that case, the votes received through the answers might be a better measure of the time and effort spent

	v_{is}	v_{is}	x_{is}	x_{is}
NewJob(s=1)	-0.198*** (-3.59)	-0.198*** (-5.24)	-0.0587 (-1.21)	-0.0587** (-2.12)
Votes(x=v)			0.433*** (8.91)	0.433*** (13.44)
NewJob×Votes			-0.140** (-2.03)	-0.140*** (-4.95)
Constant	1.004*** (25.70)	1.004*** (52.96)	0.571*** (16.62)	0.571*** (29.09)
<i>Model</i>	OLS	FE	OLS	FE
<i>N</i>	1560	1560	3120	3120
<i>R</i> ²	0.00822	0.0340	0.0400	0.0954

t statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table 4: Difference-in-Differences Estimation: Quantity and Quality of Answer Activity

on answer activity. Therefore, we conduct another set of Difference-in-Differences regression using votes as the measure of answer activity. We count all the votes received from an answer within 30 days after the answer was given. Then we add up the votes from all the answers provided by user i in period t . Votes gives more weights to answers with high-quality (more votes received), and it takes into consideration of both the quantity and quality of answers.

$$\begin{array}{l|l}
 v_0 & \log \# \text{ votes to answers in months } -4, -3, -2 \text{ given up to} \\
 & \text{one month after answer is given (monthly average)} \\
 v_1 & \text{ditto for months } 1, 2, 3
 \end{array}$$

Table 3: Additional Notation for Regression 3 and 4

$$v_{is} = \beta_1 + \beta_2 s + \epsilon_{is} \quad (3)$$

$$x_{is} = \beta_1 + \beta_2 s + \beta_3 \mathbf{1}(x = v) + \beta_4 s \mathbf{1}(x = v) + \epsilon_{is} \quad (4)$$

where

$$i \in \{1, \dots, 780\} \quad \text{as of December 15, 2014}$$

$$s \in \{0, 1\}$$

$$x \in \{v, e\}$$

Table (4) summarizes regression results using total votes as the measure of answer activity. All the regressions produce similar estimates to results in table (2) which uses the quantity of answers. The total votes gained from answers drops by 20.3%, in which 5.8 percentage points drop is due to change in time availability, and 14.5 percentage points drop is due to career concerns.

The almost identical estimates from the two measures of answer activity are not completely surprising, since the total votes sums up the votes from each answer. Another possible reason is that more efforts only contribute through the quantity but not the quality of the answers provided. In the scenario where all contributors provide their best answer given their ability, the marginal cost of quality improvement might be too high compared to the marginal gains of votes.

7 Robustness Checks

The main concern for our identification strategy is a mismatch between a job candidate's expertise and new job requirements, that is to say, the skills required for a new job might be different from those required for the previous job. In that case, a contributor might spend more time asking questions instead of contributing answers. Then the parallel trend assumption for Difference-in-Differences approach is violated. Using the detailed information available on Stack Overflow, we show several pieces of evidence in support of our identification assumption.⁵

First, tag information associated to each job allow us to conduct a triple-difference regression to compare the impact of career concerns on contribution between users whose new jobs have similar and different tags. We define a measure of similarity of skills between jobs. Then we divided the contributors into two groups: 1) Those whose new job requires similar skills to the previous job; 2) Those whose new job requires different skills from the previous job. The final triple-difference regression gives no significant differences in the impact of career concerns between these two groups.

Second, if there is significant learning effects with a new job, then we should observe a significant increase in question activity on Stack Overflow. To test whether this is the case, we conduct Difference-in-Differences regression using question and edit activities, and found that there is no significant change of question activity after a job change.

Third, if the reduction in contribution level is mainly due to a mismatch of skills, then users who ask questions in the same category should not be affected by a job change. We collect the tags information related to each question for all users, and conduct the same set of Difference-in-Differences regressions as in the previous

⁵For all the details of robustness checks, please refer to the Online Appendix at <http://leixu.org/>.

section using users whose questions are associated to similar tags based on several measures. The results are very similar to those from the previous section.

8 Conclusion

For public goods that require voluntary contribution, it is imperative to understand the sources of contributors' motivation as well as the impact on their activity. A not-well-thought-of program mechanism often fail to attract users and discourage user contributions,⁶ whereas a well-designed incentive system can significantly contribute social welfare.⁷

Education and work experience have played an important role in signaling job candidates' ability. In the digital age, recruiters can gather more signals through one's online activity. Stack Overflow provides an ideal platform through which users can learn, help each other, and at the same time signal their ability to potential employers through their participation.

By taking advantage of various types of activities on Stack Overflow, we empirically separate out extrinsic motivation and estimate the impact of career concerns on user activity. Our result shows that career concerns indeed play a significant role in their online activity. To the best of our knowledge, this is the first empirical study that measures the impact of career concerns on contributions to public goods.

9 Limitation

The current results may be subject to a selection bias, since we only focus on people who have a profile on Stack Overflow Careers, which might not be representative of all users on Stack Overflow. At the same time, there might be other factors that could potentially affect the propensity to contribute. One potential solution is to use a two-stage Heckman selection model, which was used to resolve the selection bias in similar settings.

Although we are able to assess the changes in participation due to a change in career concerns using Difference-in-Differences approach, we can not rule out the possibility that the changes may be due to an interaction between career concerns and other motives.

In the end, our research raises more questions than it answers. A lack of data on earnings and job matching quality makes it impossible to measure the effectiveness

⁶Well-known examples include include Yahoo! Answers, Digg, etc.

⁷Successful examples include Wikipedia, YouTube, Stack Overflow and Amazon Mechanical Turk.

of signaling through online contribution. Moreover, we are unsure whether this type of extrinsic motivation crowds out intrinsic ones in the long run.

10 Appendix

Proof of Proposition 1. Consider a user in state $s = 1$, an absorbing state. Since the value of tasks u_τ is ordered, an optimal policy is to select minimum thresholds s_τ of type τ tasks to undertake. The user's maximization problem is therefore given by

$$\max_{s_w, s_e, s_a} U = m_w \int_{s_w}^{\infty} \alpha dF_1(\alpha) + m_e \int_{s_e}^{\infty} \alpha dG_1(\alpha) + m_a \int_{s_a}^{\infty} \alpha dG_1(\alpha)$$

subject to

$$x_w + x_e + x_a = T$$

where

$$x_w \equiv m_w(1 - F_1(s_w))$$

$$x_m \equiv m_e(1 - G_1(s_e))$$

$$x_a \equiv m_a(1 - G_1(s_a))$$

The first-order conditions for utility maximization are given by

$$\begin{aligned} -f_1(s_w) s_w m_w &= -f_1(s_w) m_w \lambda \\ -g_1(s_e) s_e m_e &= -g_1(s_e) m_e \lambda \\ -g_1(s_a) s_a m_a &= -g_1(s_a) m_a \lambda \end{aligned}$$

where λ is the Lagrange multiplier of the time constraint. This implies $s_w = s_e = s_a$. Moreover, $x_m/m_e = x_a/m_a$. Finally, let U^* be the optimal level of utility.

Consider now a user in state $s = 0$. The maximization problem is given by

$$\begin{aligned} \max_{s_w, s_e, s_a} U &= m_w \int_{s_w}^{\infty} \alpha dF_1(\alpha) + m_e \int_{s_e}^{\infty} \alpha dG_1(\alpha) + m_a \int_{s_a}^{\infty} \alpha dG_1(\alpha) \\ &+ \delta (1 - p(r)) V_0 + \delta p(r) \frac{U^*}{1 - \delta} \end{aligned}$$

where V_0 is the continuation value if the state remains $s = 0$, $V_1 = \frac{U^*}{1 - \delta}$ is the continuation value if the state switches to $s = 1$, and $p(r) = p(x_a)$ is the probability that the state switches from $s = 0$ to $s = 1$.

The first-order conditions for utility maximization are now given by

$$\begin{aligned} -f_1(s_w) s_w m_w &= -f_1(s_w) m_w \lambda \\ -g_1(s_e) s_e m_e &= -g_1(s_e) m_e \lambda \\ -g_1(s_a) s_a m_a - g_1(s_a) p'(x_a) \delta (V_1 - V_0) &= -g_1(s_a) m_a \lambda \end{aligned}$$

Since $F_1(\cdot)$ dominates $F_0(\cdot)$ in the sense of first-order stochastic dominance, it follows that $V_1 > V_0$. This in turn implies that $x_e/m_e < x_a/m_a$ if and only if $p'(\cdot) > 0$. \square

References

- Bitzer, Jürgen, and Ingo Geishecker.** 2010. “Who contributes voluntarily to OSS? An investigation among German IT employees.” *Research Policy*, 39(1): 165–172.
- Blatter, Marc, and Andras Niedermayer.** 2008. “Informational hold-up, disclosure policy, and career concerns on the example of open source software development.” *Disclosure Policy, and Career Concerns on the Example of Open Source Software Development (September 1, 2008) .NET Institute Working Paper*, , (08-06).
- Chevalier, Judith, and Glenn Ellison.** 1999. “Career concerns of mutual fund managers.” *The Quarterly Journal of Economics*, 114(2): 389–432.
- Hann, Il-Horn, Jeffrey A Roberts, and Sandra A Slaughter.** 2013. “All Are Not Equal: An Examination of the Economic Returns to Different Forms of Participation in Open Source Software Communities.” *Information Systems Research*, 24(3): 520–538.
- Holmström, Bengt.** 1999. “Managerial Incentive Problems: A Dynamic Perspective.” *The Review of Economic Studies*, 66(1): 169–182.
- Kolstad, Jonathan T.** 2013. “Information and Quality When Motivation is Intrinsic: Evidence from Surgeon Report Cards.” *American Economic Review*, 103(7): 2875–2910.
- Lerner, Josh, and Jean Tirole.** 2001. “The open source movement: Key research questions.” *European Economic Review*, 45(4): 819–826.
- Lerner, Josh, and Jean Tirole.** 2002. “Some simple economics of open source.” *The Journal of Industrial Economics*, 50(2): 197–234.

- Milgrom, Paul R.** 1981. “Good News and Bad News: Representation Theorems and Applications.” *The Bell Journal of Economics*, 12(2): 380–391.
- Roberts, Jeffrey A, Il-Horn Hann, and Sandra A Slaughter.** 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.
- Spiegel, Yossi.** 2009. “The incentive to participate in open source projects: a signaling approach.” *Working Paper*.
- von Krogh, Georg, Stefan Haefliger, Sebastian Spaeth, and Martin W Wallin.** 2012. “Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development.” *MIS Quarterly*, 36(2).
- Zhang, Xiaoquan, and Feng Zhu.** 2011. “Group size and incentives to contribute: A natural experiment at Chinese Wikipedia.” *American Economic Review*, 101(4): 1601–1615.