What Makes Geeks Tick? A Study of Stack Overflow Careers

Lei Xu∗, Tingting Nian†, and Luis Cabral‡

1McGill University
2NYU Stern School of Business
3NYU Stern School of Business

November 27, 2014

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 the time availability which can also cause changes in contribution levels. When changing to a new job, contributors on average experience a 18.5% drop in contribution activity and 80% of the drop is due to career concerns.

PRELIMINARY AND INCOMPLETE PLEASE DO NOT CIRCULATE!

∗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 contributor's CVs. With contributors' profile on Stack Overflow Careers and their historical activities on Stack Overflow, we are able to construct complete histories of each individual's online trajectory: contributions to Stack Overflow as well as inclusion date and employment histories on Stack Overflow Careers.

The goal of the 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 and Stack Overflow.

2 Theoretical Model

We propose a simple dynamic model of user contributions. Consider a discrete time, infinite period model with discount factor $\delta$. Each user has a measure of time $T$ to distribute between three types of tasks: $\tau = w, e, a$, where $\tau$ 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 \( \tau \) tasks available is given by \( m_\tau \). 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. Moreover, 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

**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.

### 3 Identification Strategy

Our identification relies on a reduction of career concerns when a contributor changes jobs. A comparison between online activities before and after the job change can reveal the effect of career concerns.

However, a confounding factor that might cause a reduction in online activity is the time availability. So we conduct difference-in-differences regressions using Edits as a comparison group which proxies for time availability.

Votes can be partly caused by extrinsic motivation (e.g. career concerns), whereas Edits are caused only by intrinsic motivation (e.g. help the community). Therefore, Edits could not be affected by any extrinsic motivation but could be affected by the time availability.

Without a change in career concerns, a change in time availability should lead to a parallel change in both types of activities. So the first difference measures the change of contribution level before and after the job change for both types of activities. For example, a contributor changes his or her job in July, 2014, then we measure the changes of the number of edits and votes in several months before and after July. The second difference is between these two types of activities. The Diff-in-Diff result measures the change in contribution level due to career concerns but not due to the time availability.
The identification strategy can be illustrated in Figure 1. With a job change, the reduction of Votes comes from two sources: 1) Career concerns 2) Time availability. At the same time, the reduction of edits has only one source which is the time availability. A preliminary plotting of average activities on Stack Overflow (Figure 2) shows the validity of our identification strategy.

Figure 2 shows that there is a significant reduction in Votes but not Edits. At the same time, the trends of Votes and Edits before and after the job change seem parallel too. Next section will further test the validity of the common trend assumption in Diff-in-Diff regression.

4 Regression Analysis

4.1 Job Changes

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

Our hypothesis is that due to career concerns, a job candidate works harder in order to improve her signal to employers through contribution activities on Stack Overflow. If this is the case, then after a job change, the effect of career concerns diminishes and the contribution level drops.

To evaluate their hypothesis, we collect a panel of contributors’ job changing dates and
their associated Stack Overflow IDs. With their IDs, we collect their contribution activity from four-month before and four-month after the date of job change. Then we use the following difference-in-differences regression to test the impact of changing jobs on the contribution level:

\[
\text{Activity}_{ijt} = \beta_1 + \beta_2 1(\text{NewJob}_t) + \beta_3 1(\text{Votes}_j) + \beta_4 1(\text{NewJob}_t) \times 1(\text{Votes}_j) + \epsilon_{ijt} \tag{1}
\]

The dependent variable, \(\text{Activity}_{ijt}\), includes two types of activities: Answers \((j = 0)\) and Edits \((j = 1)\). By answering questions, contributors gain reputation points through votes received. \(\text{Activity}_{i0t}\) calculates the total votes gained from answers provided by user \(i\) at time \(t\); \(\text{Activity}_{i1t}\) counts the total edits to others’ posts contributed by user \(i\) at time \(t\). \(1(\text{NewJob}_t)\) is a dummy variable which indicates whether time \(t\) is after a job change. \(1(\text{Votes}_j)\) indicates whether the value of the dependent variable \(\text{Activity}_{ijt}\) represents Answers or Edits.

Table II summarizes the regression results using both OLS and fixed-effect regression. The fixed-effect regressions enable us to focus on the change in each individual’s contribution level before and after a change in job seeking status while accounting for time-invariant individual characteristics.

Model 1 and 2 measure the average answers activity level for contributors before and after a job change. Both OLS and FE regressions shows a significant drop of 18.5% in answers activity after a job change.

\(^2\)In the future, we will also adopt other measures of contribution level, including number of answers and word count of answers.
Table 1: Difference-in-Differences Estimations of the Impact of Career Concerns on Contributors Level

<table>
<thead>
<tr>
<th>Model</th>
<th>OLS FE</th>
<th>OLS FE</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>2804</td>
<td>2804</td>
</tr>
<tr>
<td>R²</td>
<td>0.010</td>
<td>0.042</td>
</tr>
</tbody>
</table>

Model 3 and 4 conduct Difference-in-Differences regressions. After removing the effect due to a change in time availability proxied by the number of edits, there is a significant drop of 14.8% in votes-generating activity. That is to say, within the 18.5% reduction in total votes from answers, 14.8% is due to career concerns, and 3.7% is due to the change in time availability.

5 Conclusion

For online communities such as Stack Overflow which are trying to attract new users and sustain existing users to participate in the voluntary contribution with their knowledge and expertise, it is imperative to understand the relationship between different types of motivations and their participation level. Since Stack Overflow, like many Open Source Software (OSS) communities, offers a great setting for users to signal their abilities through participation, career concerns may be one of the most salient drivers. In this paper, we empirically separate the impact of career concerns from other types of motivations. Using users’ job change history as an approximation for their changing underlying career concerns, our paper elucidates the substance and magnitude of the impact of career concerns on the participants’ contribution. To our knowledge, there is the first study that finds empirical evidence of the effect of career concerns on users’ online contributions.
References


