

What Makes Geeks Tick?

A Study of Stack Overflow Careers

Lei Xu

McGill University

Tingting Nian

New York University and UC Irvine

Luís Cabral

New York University and CEPR

Preliminary and incomplete: May 2015

Abstract. We study user contributions to Stack Overflow, the largest online programmers community. Using a difference-in-difference approach, we show that the event of finding a new job implies a reduction of 25% in reputation-generating activity, but only a reduction of 8% in non-reputation-generating activity. We consider a series of robustness tests to tease out alternative explanations for these variations; together, the results suggest that career concerns play an important role in explaining user contributions.

Keywords: career concerns, signaling, online contribution, intrinsic and extrinsic motivation

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

Xu: Visiting Scholar, Business Economics, Rotman School of Management, University of Toronto; Ph.D. Candidate, McGill University; lxu@stern.nyu.edu. Nian: Department of Information, Operations & Management Sciences, NYU Stern School of Business; The Paul Merage School of Business at UC Irvine; tnian@stern.nyu.edu. Cabral: Paganelli-Bull Professor of Economics and International Business, Stern School of Business, New York University; luis.cabral@nyu.edu

We thank Allan Collard-Wexler, Bryan Bollinger, Kei Kawai, Ariel Pakes, Jidong Zhou, Ignatius Horstmann, Heski Bar-Isaac, Philip Oreopoulos, Fabian Lange, Yan Song, Avi Goldfarb, Mitchell Hoffman, Andrew Ching, Victor Aguirregabiria, Aloysius Siow for valuable suggestions; and seminar participants at NYU Stern, McGill, UToronto, UC Irvine (Merage), University of Washington (Foster), University of Rochester (Simon), HKUST Business School, Singapore Management University and WISE 2014 for helpful comments. We acknowledge support from the McGill University, NYU Stern, and Rotman. All errors are our own.

1. Introduction

The Internet has revolutionized the world in more than one way. Of particular interest is the phenomenon of private contributions to collective projects such as Wikipedia, bulletin boards, or open source software. As Lerner and Tirole (2002) put it, to an economist the behavior of individual contributors appears somewhat puzzling: is it a case of altruism, or are there ulterior motives behind private contributions to a public good?¹

Our paper addresses this research question using data from Stack Overflow, one of the leading 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: the desire to signal one’s ability so as to obtain better employment.

Associated to Stack Overflow (SO), the Stack Overflow Careers (SOC) site hosts job listings and contributors’ CVs so as to match employers and employees. The information regarding each job candidate includes their employment history as well as various summary statistics regarding their contribution to SO. A tantalizing possibility — the hypothesis we propose to test — is that contributing to SO is a way of signaling one’s ability and thus find a better job.

We construct complete histories of each individual’s online trajectory. This includes the contribution to SO as well as individual characteristics and employment history. We test the career-concerns hypothesis by identifying shifts in behavior following career-relevant shifts, namely employment changes.

We find that before changing to a new job, a contributor gives more and better answers but asks fewer questions. However, right after the job change, there is a significant drop in answers both in terms of quality and quantity. We conclude that contribution levels decrease by 22.3% right after a job change, and 14.5% is due to change in the level of career concerns.

To the best of our knowledge, ours is the first paper empirically to estimate the relation between changes in career status and free contributions to online communities as an indirect measure of career concerns. We believe our methodology can be helpful in other contexts; and we believe our empirical results are important, considering the increasing prevalence of user-generated content in a variety of settings (Wikipedia, Stack Overflow, GitHub, YouTube, etc).

■ **Related literature.** We are by no means the first to examine empirically the phenomenon of free user contributions. Zhang and Zhu (2011) show that, when access to Chinese Wikipedia was blocked in mainland China in October 2005, contributions by users outside of mainland China dropped by more than 42.8%, an effect that they attribute to the intrinsic and internalized-extrinsic motivation for free user contribu-

1. von Krogh et al. (2012) distinguish three types of motivation: intrinsic (e.g., altruism, ideology, fun, kinship); internalized extrinsic (e.g., reputation, learning, reciprocity, own-use); and extrinsic (e.g., career concerns, pay).

tions (see von Krogh et al. (2012)). By contrast, we show that extrinsic motivations also play an important role.

Most of the theoretical and empirical literature on free user contributions addresses the issue of Open Source Software (OSS). In many ways, the OSS phenomenon is very similar to free user contributions to sites such as SO or Wikipedia, so a review of this literature is warranted.² At a conceptual level, Lerner and Tirole (2001), Blatter and Niedermayer (2008) and Mehra, Dewan, and Freimer (2011) show how contributors to OSS projects may lead to better career prospects. At an empirical level, Bitzer and Geishecker (2010) show that the propensity to work on OSS projects is higher among university dropouts, a pattern which they interpret as evidence of career-oriented motivations. Roberts, Hann, and Slaughter (2006) and Hann, Roberts, and Slaughter (2013) find evidence of subsequent returns from participation in OSS projects.

2. Theoretical model

Consider an infinite-period, discrete time line, and suppose agents discount the future according to the factor δ . Each agent is a SO contributor and a job seeker. The agent's state space is limited to $s \in \{0, 1\}$, where $s = 0$ stands for current (or old) job and $s = 1$ stands for future (or new) job. We assume $s = 1$ is an absorbing state. To the extent this is not the case, our estimates of career concerns should be regarded as a lower bound of the real size of career concerns.

A fundamental hypothesis that we propose to test is that the probability of job transition — that is, the transition from $s = 0$ to $s = 1$ — is endogenous, specifically, a function of the agent's reputation:

$$\mathbb{P}(s_t = 1 \mid s_{t-1} = 0) = p(r_t)$$

In each period, agents must decide how to allocate their time. We consider three types of tasks: Work, Answers and Edits. Let w_t, e_t and a_t be the time devoted to each of these tasks. Each agent's time constraint is then given by

$$w_t + e_t + a_t = T$$

Consistently with the structure of SO, we assume that r_t is a function of past values of a_t *but not* of past values of e_t . In fact, a crucial difference between Answers and Edits is that the former is a vote-generating activity whereas the latter is not.³

We assume each agent's utility each period is additively separable in each of the three tasks:

$$u_t = g_s(w_t) + f(e_t) + f(a_t) \tag{1}$$

2. Spiegel (2009) highlights the theoretical difference between free contributions to OSS and to Stack Overflow (whereas the former might succeed or fail, users always benefit from higher contribution levels in the latter).

3. In addition to Answers, Questions are also a vote-generating activity. For simplicity, we limit our theoretical analysis to the case of one vote-generating activity. In the empirical part of the paper we also consider Questions as part of an agent's optimization process.

where $f(\cdot)$ and $g(\cdot)$ are twice differentiable functions such that $f', g' > 0$ and $f'', g'' < 0$. Notice we allow the utility from work to be state-dependent. In fact, the agent's demand for a new job results from our assumption that $g_1(w) > g_0(w)$.

Agents are forward looking: in each period, they choose w_t, e_t, a_t so as to maximize value V_s , where $s = 0, 1$. The value functions are determined recursively as follows:

$$\begin{aligned} V_0 &= g_0(w_t) + f(e_t) + f(a_t) + \delta p(r_{t+1}) V_1 + \delta (1 - p(r_{t+1})) V_0 \\ V_1 &= g_1(w_t) + f(e_t) + f(a_t) + \delta V_1 \end{aligned}$$

Finally, we assume reputation at time t is given by the number of answers in $t - 1$

$$r_t = a_{t-1}$$

(In the empirical part of the paper we consider various other possibilities. The qualitative nature of our theoretical results remains valid if we assume more complicated reputation functions.)

Our main theoretical result is that a change in state (getting a “better” job) implies an absolute decline in the number of answers as well as a decline in the ratio answers / edits. Moreover, the latter takes place if and only if career concerns matter:

Proposition 1. *Suppose that $g_0(w) < g_1(w)$. Then*

$$a_t \Big|_{s=1} < a_t \Big|_{s=0}$$

Moreover,

$$\frac{a_t}{e_t} \Big|_{s=1} < \frac{a_t}{e_t} \Big|_{s=0} \quad \text{iff} \quad p'(\cdot) > 0$$

Proof of Proposition 1: Let x_s^* be the optimal value of control variable x ($x = w, e, a$) in state s ($s = 0, 1$). Suppose that $V_1 \leq V_0$. Then by choosing $x_t = x_0^*$ when $s = 1$ a strictly higher value of V_1 and V_0 is obtained. Thus it must be

$$V_1 > V_0 \tag{2}$$

At state s , the agent maximizes V_s subject to $w + e + a = T$ (for simplicity we omit the time subscript). The first-order conditions at $s = 1$ are given by

$$\begin{aligned} \lambda_1 &= g_1'(w_1) \\ \lambda_1 &= f'(e_1) \\ \lambda_1 &= f'(a_1) \end{aligned} \tag{3}$$

where λ_1 is the Lagrange multiplier in state 1. At $s = 0$, we have

$$\begin{aligned} \lambda_0 &= g_0'(w_0) \\ \lambda_0 &= f'(e_0) \\ \lambda_0 &= f'(a_0) + \delta p'(a_0) (V_1 - V_0) \end{aligned} \tag{4}$$

The second and third equations in (3) imply that $e_1^* = a_1^*$. The second and third equations in (4), together with (2), imply that $e_1^* < a_1^*$ if and only if $p'(a_0) \neq 0$. Together, these equations imply the second part of the Proposition.

Regarding the first part of the Proposition, it helps to make the comparison in two steps. First consider changing (3) by substituting g'_0 for g'_1 . Since, $g'_1 > g'_0$, this results in a higher value of a (and of e). Second, consider adding the term $\delta p'(a_0) (V_1 - V_0)$. Given (2) and to the extent that $p' \geq 0$, again the value of a increases. ■

Proposition 1 establishes two effects of a job change: a decline in the time spent on Answers; and a decline in the relative time spent on Answers vis-a-vis Edits. The first effect (decline of Answers) can be decomposed into two effects: an increase in the marginal utility of time spend at work; and a decline in the utility of Answers derived from career concerns. Since there are two effects, a decline in Answers is a necessary but not sufficient condition for our career-concerns hypothesis. By contrast, the second effect takes place if and only if career concerns are present. It provides, therefore, a sharper test of our central hypothesis.

One advantage of a theoretical model is that it helps clarify the assumptions underlying an empirical identification strategy. The assumption that the Edits and the Answers components in the utility function share the same functional form $f(\cdot)$ plays an important role. It can be shown that the results go through if these components are the same up to a linear transformation. This is an important point because, if taken literally, our model implies that $e_t = a_t$ while in state $s = 1$, a very strong restriction. Allowing for the Edits and Answers components in the utility function to vary gives an extra degree of freedom regarding levels while maintaining the result regarding the relative values of e and a . For example, one possible functional form would be

$$u_t = \alpha w_t^{\beta_s} e_t^\gamma a_t^\eta$$

Taking logarithms (a monotonic transformation of the utility function, which therefore does not change optimal choices), we get an expression similar to (1) except that the coefficients γ and η appear in front of $f(\cdot)$, which in the present case is given by $f(\cdot) = \ln(\cdot)$.

3. Data

Our dataset is derived from the Stack Overflow (SO) and Stack Overflow Careers (SOC) sites. SO is the largest programming site where programmers ask and answer programming-related questions. It provides for Wikipedia-style editing; and it includes a system of votes, badges and user reputation that ensures high-quality, peer-reviewed answers. SO is widely used by both programmers and programming-related companies. Founded in 2008 by Joel Spolsky and Jeff Atwood, it currently comprise 3.5 million users. Some summary statistics regarding the site's activity: 6.7 million visits/day; 7.4 thousand questions/day; 8.1 million cumulative questions, 14 million cumulative answers.

SOC is a related job matching website that hosts programming-related job listings as well as candidate resumes. For contributors, creating a resume on the website (as shown in Figure 1) is free of charge but by invitation only; and the invitation is based on the contributors' recent activity to the site as well as their field expertise.⁴ On the resume, contributors can easily provide a link to their SO profile, through which employers can learn more about the job applicants' expertise: that is, potential employers observe the user's reputation score, a reflection of the quantity and quality of the user's contribution to SO.

SOC helps employers by reducing hiring search costs (although access to SOC is paid). First, SOC provides a select sample of high-level contributors invited by SO. Second, SOC includes a wealth of information regarding the job applicants' skill sets, including in particular their contribution history to SO. Finally, employers who access SOC may post their openings as well as search candidates by location, skills, and so on.

■ **Measures of user activity.** We divide a user's activity on SO into four different categories:

- Questions Any registered user can ask a Question. A Question can be voted up or down. A hard but important Question is usually voted up to get attention from more contributors. A duplicate or unclear Question is usually voted down.
- Answers Any registered user can provide Answers to others' Questions. A Question can have multiple Answers and the latter are ranked by total votes.
- Edits Registered users can also make or suggest minor changes to questions and answers: Edits.⁵ Edits help make the questions and answers more readable and understandable to future viewers.
- Votes Finally, registered users can give upvotes or downvotes to Questions and Answers *but not to Edits*. Votes contribute to the post owner's reputation: upvotes on a Question give the asker 5 points, whereas upvotes on an Answer are worth 10 points.

■ **Data selection.** We focus on a set of users that satisfy a series of criteria required by our empirical test:

- Located in the U.S. and Canada: this ensures a more homogenous sample.⁶

4. The exact criteria is not disclosed by SO. An alternative path to an invitation is to request it on the website.

5. Most Edits correct grammar or spelling mistakes; clarify the meaning of a post; or add related information.

6. A large fraction of the jobs posted on SOC are located in the United States and Canada.

Figure 1
Sample profile on Stack Overflow Careers

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`

Table 1
Descriptive statistics

	Mean	Median	Std. Dev.	Min.	Max.
User Activity					
Answers	4.055	0	12.310	0	417
Votes (from Answers)	5.967	0	23.023	0	966
Questions	0.637	0	1.933	0	58
Edits	1.748	0	9.883	0	689
User Characteristics					
Profile Views	359.723	71	2170.283	0	112,967
Total UpVotes	334.669	82	800.728	0	15,143
Reputation Points	1603.965	150	6204.839	-6	132,122
Age	33.889	33	7.433	16	95
Time on SO	4.225	4.337	1.503	0.167	6.507

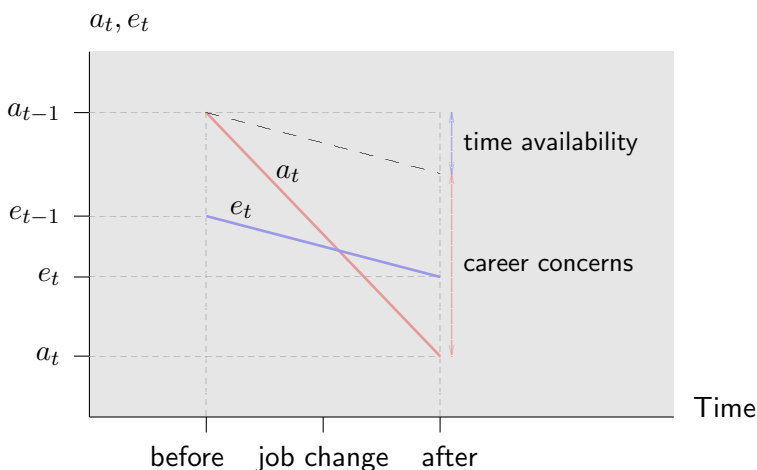
- Job changers: the change in the level of career concerns comes from a job change; we select users who experienced a job change from November 2008 until November 2014, the month when we stopped collecting data.
- Job switchers: Employment status (employed vs. unemployed) introduces unnecessary noise; we select users who *switch* from one job to another (with a gap less than or equal to one month between two jobs).⁷
- Active users: for many users, we do not observe any activity on SO during periods of job change; for more accurate estimation, we focus on active users, defined as having at least one answer and at least one edit within four months before and after the month of job change.
- Profile with link to SO: the ability to track users' online activity requires the link to SO.

Applying this series of criteria results in a sample of 1249 users with 1500 job switches. Obviously, the sample we use is not representative of the population, so coefficient estimates should be interpreted accordingly. We return to this issue later.

Table 1 provides some descriptive statistics. Looking at the user activity, we see that the typical SO user is not active in writing Questions or Answers — or voting, for that matter. The activity distributions are fairly right-skewed, suggesting that a few users are disproportionately responsible for much of the content created on SO.

7. We are currently also working on the analysis of contributors who experience a change in employment status, i.e., from unemployed to employed.

Figure 2
Identification: difference-in-differences



The lower portion of the table suggests that typical users of SO are in their early 30s and have been on SO for 4 years (SO has existed for 7 years).

4. Identification strategy

Conceptually, our identification strategy is quite straightforward: job seekers are active on SO to signal their ability and thus obtain a better job. If career concerns are important, then we expect a drop in such activity once the goal (a better job) is attained. Since no one expects to remain in the same job for the rest of their lives, career concerns might not entirely disappear; but at least they are diminished following a change in jobs.

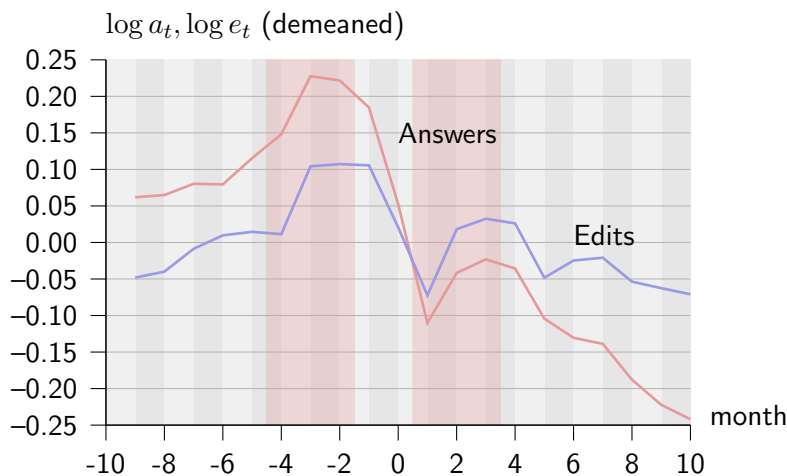
In practice, there are various confounding factors that make measurement of career-concern effects difficult. In particular, a reduction in online activity following a job change may simply result from a reduction in time availability: a new job often requires training and more generally some time investment so as to be familiarized with a new environment. In fact, as the first part of Proposition 1 states, we expect a drop in a_t through two effects: a drop in career concerns (measured by $p(r_t)$ in the model); and an increase in work activities (measured by the shift from $g_0(w)$ to $g_1(w)$ in the model).

To account for these effects, we conduct difference-in-differences regressions using Edits as a control group that proxies for time availability. A crucial difference between Edits and Answers is that the latter give rise to votes, whereas the former does not. Therefore, we expect the career-concerns effect to act through the Answers channel but not through the Edits channel. (Questions are also reputation generating activities; we will focus on these later.)

The implicit assumption in our difference-in-difference approach is that, aside from changes in job status, Edits and Answers follow a parallel path. Essentially, this

Figure 3

Average monthly activity by period



corresponds to the assumption in Section 2 regarding the utility function functional form. Since this is such a crucial assumption, in Section 6 we provide various pieces of evidence in its support.

Essentially, our difference-in-difference approach corresponds to the second part of Proposition 1. Figure 2 illustrates the main idea: after starting a new job, the reduction in Answers activity results from two effects: career concerns and time availability (or, opportunity cost of work time); however, the reduction in Edits activity results exclusively from the time availability effect; therefore, the difference between the changes in Answers and in Edits identifies the effect of job change on career-concerns incentives for Answers.

Figure 3 provides preliminary evidence regarding our hypothesis. It plots the monthly average of the logarithm of user activity in a 20-month window centered around a contributor's job change event. As can be seen, both Answers and Edits activity experience a significant drop when a user starts a new job (month 1); however, the drop in Answers activity is considerably more significant than the drop in Edits activity. This evidence is consistent with the hypothesis that $p'(a_t) > 0$, that is, an increase in Answers increases the probability of a job change.

Naturally, several other factors may explain these dynamics. In the next section, we carry our test one step forward by running a series of regressions that explain the variation in Answers activity.

5. Regression analysis

We now come to a more formal test of the hypothesis implied by Proposition 1. Our test is based on a panel of users who changed their job status from November 2008 to November 2014. We associate user resumes (which include dates of job changes) to the user's SO ID. With the user ID at hand, we then collect the user's activity

during the four-month period before and after the month of job change: number of Answers and number of Edits (later we consider other activity measures).

Our sample is based on a series of criteria. First, so as to get some uniformity in the type of users, we restrict attention to SOC members who are located in the US. From this set, we exclude users whose resumes do not include a link to SO (and for which we lack the crucial activity data). We also exclude users who did not change jobs during our focus period. Some users experienced more than one switch; we exclude such switches if they are less than 8 months apart. Finally, we exclude users who, during the period under analysis, had a zero level of activity (in other words, we exclude inactive SO users).

As the result of this sample selection criteria, we end up with 1,249 users who were subject to 1,500 job switches during the November 2008–November 2014 period. For each of these job switches, we measure activity levels by activity type and by month. Specifically, define month 0 as the month when the job change took effect (that is, the month listed on the resume as starting month for the new job). We then consider 3 months prior to a job switch ($-4, -3, -2$); and 3 months subsequent to a job switch ($+1, +2, +3$). We thus exclude months -1 and 0 ; in this way we get a cleaner perspective on the periods before and after the job change without contaminating the data with noise stemming from the process of job change.

We then use the following regression specifications to estimate the impact of career incentive on online voluntary contribution:

$$y_{it} = \alpha_i + \beta s + \epsilon_{is} \tag{5}$$

$$y_{it} = \alpha_i + \beta s + \gamma d s + \epsilon_{is} \tag{6}$$

In these equations, y represents a generic activity: $y = e, a$; as before, s is the state: $s = 0$ corresponds to the period before a job change takes place, whereas $s = 1$ corresponds to the period after a job change takes place; α, β and γ are parameters to estimate; and ϵ represents the equation’s residual. Finally, d is a dummy variable that takes on the value 1 if the activity in question is a vote-generating activity (Answers) and 0 otherwise (Edits).

Equations (5) and (6) essentially correspond to the two parts of Proposition 1. Specifically we expect the level of SO activity to drop subsequently to a job shift, that is, we expect β in (5) to be negative. Moreover, we expect the drop in Answers ($d = 1$) to be greater than that of Edits ($d = 0$), so that $\gamma < 0$ in (6), in addition to $\beta < 0$.

Table 2 shows the results for our base regressions. There are three pair of regressions, which differ in terms of regression method (fixed effects or negative binomial); and in terms of dependent variable measurement (levels, logarithms). For each pair, the first regression is limited to the activity Answers. We thus have 9,000 observations (1,500 job switches times 6 months: three prior to the job switch, three subsequent to the job switch); the second regression, includes both Answers and Edits, thus doubling the number of observations.

Consider the third pair of regressions (that is, Regressions 5 and 6), where Answers

Table 2
Explaining activity y (Answers and Edits on SO)

	(1) $y = a$	(2) $y \in \{a, e\}$	(3) $y = a$	(4) $y \in \{a, e\}$	(5) $y = a$	(6) $y \in \{a, e\}$
s	-2.079*** (0.28)	-0.466*** (0.14)	-0.312*** (0.03)	-0.149*** (0.03)	-0.252*** (0.03)	-0.080***
$s \times d$		-1.614*** (0.31)		-0.162*** (0.04)		-0.172*** (0.03)
Data	Count	Count	Count	Count	Log	Log
Regression	FE	FE	NB	NB	FE	FE
N	9000	18000	9000	18000	9000	18000
R^2	0.014	0.011			0.022	0.015

Robust standard errors in parentheses. NB: negative-binomial regression; FE: fixed-effects regression
 y : generic online activity; a : Answers; e : Edits
 $s = 1$ prior to job switch, $s = 1$ after job shift
 $d = 1$ if $y = a$, $d = 0$ if $y = e$

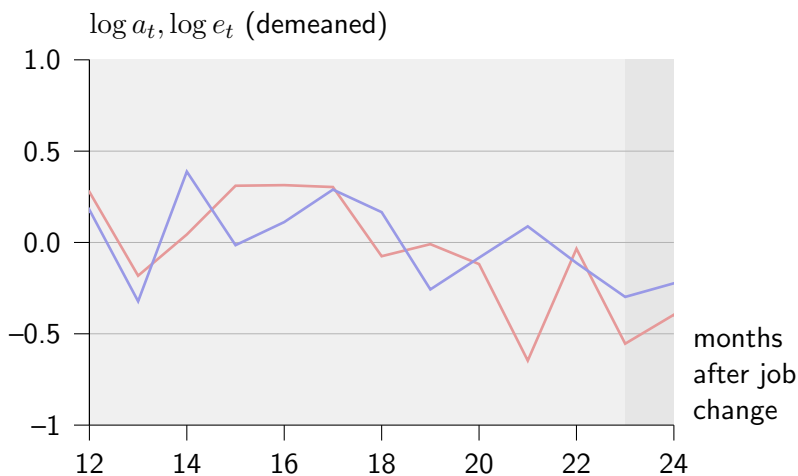
and Edits are measured in logarithms. One advantage of this approach is that the coefficients can be readily interpreted as percent variations. The fifth regression (first in the pair) has Answers as a dependent variable and s (job status) as the sole explanatory variable. The coefficient estimate on s suggests that, upon changing jobs, users decrease their Answers activity by about 25%. This is consistent with the first part of Proposition 1. However, as we mentioned earlier, it provides a weak test for our main hypothesis.

The sixth regression (second regression of the last pair of regressions) has SO activity as a dependent variable, including both Answers and Edits. As dependent variables we have s and s interacted with d , the dummy indicator of the nature of the activity. Proposition 1 predicts that both coefficients are negative. The results confirm the prediction: a job switch is associated to an 8% decline in activity (both Edits and Answers) and a further 17% reduction in Answers; this later variation we attribute to career concerns.

6. Extensions and robustness checks

In our basic test of career concerns we made a series of decisions regarding measurement of activity levels. Moreover, our identification strategy is based on a fundamental assumption, namely that the relative utility of Edits and Answers (aside from career concerns) remains constant. In this section, we consider a variety of additional results that address these and other issues.

Figure 4
Monthly Answers and Edits activity



■ **Parallel-trend assumption.** Any difference-in-difference analysis requires a parallel trend assumption, and our case is no exception. As we showed in Section 2, we require that, were it not for a job shift, the relative importance of Edits and Answers would have remained constant. Since this assumption plays a central role in our identification strategy, additional evidence on it is warranted.

For each contributor with a profile on SOC, we identify a period of time when no job change took place, that is, a period of stable employment. It seems reasonable to assume that, during these periods, the change in the level of career concerns is small compared to what we observe around the time of a job shift. Thus, consistent with our basic identifying assumption, we expect the ratio a_t/e_t to remain constant.

Figure 4 shows the values of Answers and Edits for months 12 to 24 after an agent’s job shift. Consistently with our underlying assumption, the ratio between the two is fairly constant.

■ **Expertise mismatch.** An alternative interpretation for the drop in Answers following a job shift is that the new occupation requires skills different from the previous jobs’. For example, a C++ programmer may switch to a job that is based on Java; such SO user would then be spending more time learning Java than answering C++ questions (in fact, such user might spend more time asking questions rather than answering them).

User profiles on SOC provide detailed information regarding work experience as well as user-provided information on the technology associated with each job, in the form of tags. We use tag information associated to each job to conduct a triple-

Table 3

Weekday and weekend activity on SO
 Dependent variable: Answers and Edits

	(1)	(2)	(3)	(4)
s	-0.152*** (0.03)	-0.068*** (0.02)	-0.207*** (0.08)	-0.066*** (0.02)
$s \times d$	-0.166*** (0.04)	-0.168*** (0.04)	-0.177* (0.10)	-0.131*** (0.04)
Day	Weekday	Weekday	Weekend	Weekend
Data	Count	Log	Count	Log
Regression	NB	FE	NB	FE
N	16416	16512	5076	5136
R^2		0.014		0.018

Robust standard errors in parentheses.

NB: negative-binomial regression; FE: fixed-effects regression.

$s = 1$ prior to job switch, $s = 1$ after job shift.

$d = 1$ if $y = a$, $d = 0$ if $y = e$

difference regression. First, we define a measure of skill-similarity between jobs. Then we divide our sample into two groups: high and low similarity. Finally, we re-estimate our basic regressions by adding the interaction between s , d and the dummy k , where $k = 1$ denotes high similarity across jobs. We find the triple-difference coefficient is not statistically different from zero.

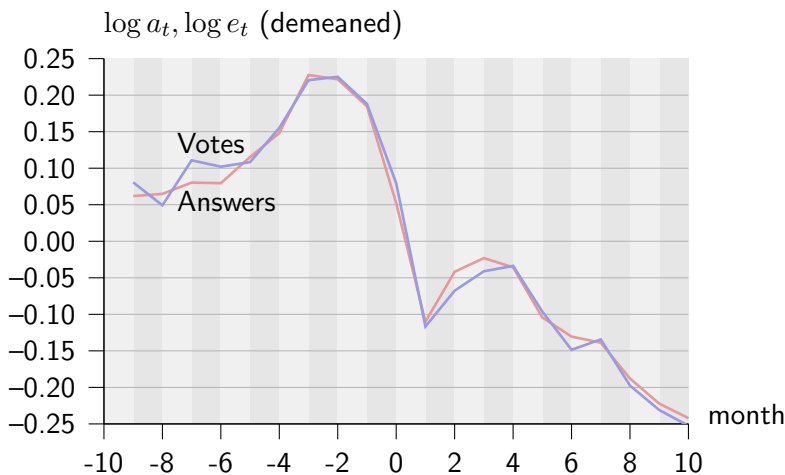
■ **Integer constraints.** One alternative interpretations for our result that a_t/e_t drops subsequent to a job change is that Answers require a bigger set-up cost than Edits. When an agent switches jobs, thus becoming busier (i.e., increasing w_t), there may be fewer time windows to justify working on an Answer rather than an Edit. In other words, Edits typically require less time and can thus be fitted into a busy schedule more easily.

To address this possibility, we split our sample into weekday and weekend SO activity. The idea is that, to the extent that work hours are more highly concentrated on weekdays, the above interpretation should imply a bigger effect on a_t/e_t during weekdays.

Table 3 shows the results of our basic regressions split into weekday and weekend days (for negative-binomial and fixed-effects regressions). Broadly speaking, the coefficient estimates are similar to those in the base model. The FE regressions show a more negative coefficient for the weekday subsample. This is consistent with the “set-up-cost” alternative hypothesis outlined above. However, the difference is rather small (about 3%).

Figure 5

Number of Answers and number of Votes to Answers



■ **Quality vs quantity measures.** Our basic results are based on quantity measures of SO activity: the number of Answers and the number of Edits. In principle, it is possible that the effect of a job shift is also felt in terms of the quality of answers. To address this possibility, we measure activity by number of Votes (given to Answers) rather than the number of Answers. Figure 5 plots the time evolution of Votes and Answers. The correlation between the two measures is remarkably high. In other words, we find no evidence of effects on the quality of answers. Not surprisingly, our tests of career concerns based on Votes rather than Answers produce very similar results (which, for brevity, we omit).

■ **Repeated changes.** Our theoretical model assumes — somewhat unrealistically — that agents change jobs at most once in their lifetime. For most people, finding a new job is not the end of a person’s career progress. In fact, in your 6 year sample period we observe a number of agents changing jobs more than once. This variation in the data allows us to push our test one step further. Let τ be the expected length of a user’s new job. Suppose that τ is observed by the agent but not by the econometrician. A reasonable proxy for τ is ex-post job tenure. Unfortunately, given the relatively narrow window we have data for (that is, given the relatively recency of SO), most of our observations are censored (we have not yet observed the next job shift). Given the data we have, a practical solution is to create a dummy variable t that takes the value 1 if a job switch is followed by another job switch in our sample. To the extent that these job switches are less permanent, a natural extension of our theoretical model predicts that the effect of a job switch is lower.

Table 4 shows the results of our regressions where we add a triple-interaction term: $s \times d \times t$. As per the discussion in the previous paragraph, we expect the coefficient to be positive, that is, to moderate the effect of $s \times d$. The results for the regression on logarithms have a lower statistical significance level. However, all coefficients have

Table 4

Repeated job changes

Dependent variable: Answers and Edits

	(1)	(2)	(3)
s	-0.427*** (0.14)	-0.154*** (0.03)	-0.080*** (0.02)
$s \times d$	-1.285*** (0.33)	-0.107*** (0.04)	-0.123* (0.04)
$s \times d \times t$	2.214*** (0.50)	0.340*** (0.06)	0.325* (0.06)
Data	Count	Count	Log
Regression	FE	NB	FE
N	17,544	17,544	17,544
R^2	.013		0.019

Robust standard errors in parentheses.

NB: negative-binomial regression; FE: fixed-effects regression.

 $s = 1$ prior to job switch, $s = 1$ after job shift. $d = 1$ if $y = a$, $d = 0$ if $y = e$ $t = 1$ if there is a subsequent job switch, $t = 0$ otherwise

the sign predicted by theory, thus reinforcing our interpretation of the changes in Answers following a job shift as resulting from changes in career incentives.

■ **Reputation size and career concerns.** Table 5 displays the results of our basic model where we split the sample by reputation size (top third, middle third, bottom third). Low-reputation users do not seem to change their behavior as a results of a job change. By contrast, medium- and high-reputation users show effects of magnitudes similar to our base regressions, with medium-reputation users a little above average and high-reputation users a little below average.

■ **Questions.** As mentioned earlier, SO users can build a reputation by answering questions but also by posting questions. Figure 6 shows the rate of Questions asked around the time of a job shift. Unlike Answers and Edits, we observe little change in the number of Questions. When we redo our basic regressions with Questions instead of Answers as a vote-generating activity, we obtain a positive estimate of the $s \times d$ interaction coefficient. Moreover, the size of the coefficient is approximately equal, in absolute value, to the coefficient on s . In words, whereas the number of Edits is reduced following a job shift, the number of Questions seems not to change. One possible explanation is that, more than a reputation-increasing activity, Questions are used as a learning tool; and a shift to a new job creates new learning demands, an effect that seems to compensate the higher opportunity cost of time spent on SO

Table 5

Regression results by reputation size
 Dependent variable: Answers and Edits

	(1)	(2)	(3)
s	0.019 (0.04)	-0.126*** (0.04)	-0.178*** (0.04)
$s \times d$	-0.083 (0.06)	-0.208*** (0.06)	-0.155** (0.07)
Reputation percentile	0-33%	34-66%	67-100%
N	1424	1424	1424
R^2	0.003	0.080	0.076

Robust standard errors in parentheses.

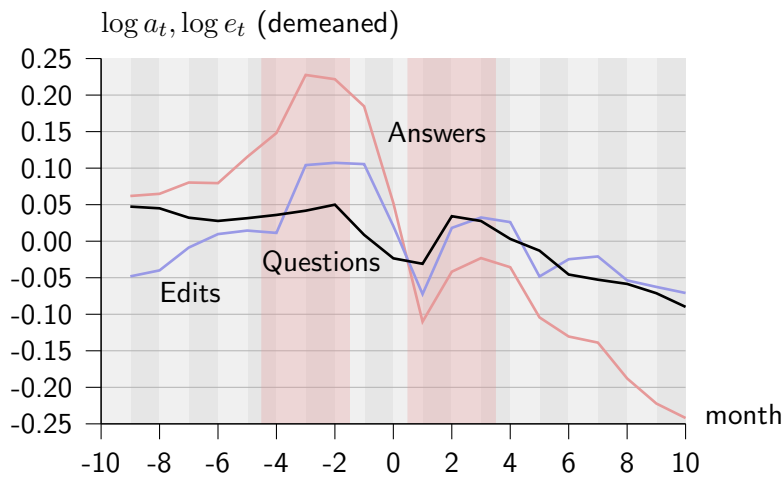
y : generic online activity; a : Answers; e : Edits

$s = 1$ prior to job switch, $s = 1$ after job shift

$d = 1$ if $y = a$, $d = 0$ if $y = e$

Figure 6

Questions, Answers and Edits



as well as the diminished incentive to build a reputation.

7. Discussion and concluding remarks

The Internet has created many opportunities for online collaboration and networking. Some platforms have been enormously successful, some less so. Examples of the former include Wikipedia, YouTube, Stack Overflow and Amazon Mechanical Turk; examples of the latter include Yahoo! Answers, Digg. What distinguishes a winner from a loser platform? We suggest that, as often is the case in economics, incentives matter, both intrinsic and extrinsic incentives. In the context of Open Source Software, Lerner and Tirole (2002) emphasize that distinguishing between these two incentive sources would “provide lenses through which the structure of open source projects, the role of contributors, and the movement’s ongoing evolution can be viewed.” The same applies, we would argue, to other types of collaboration projects as well.

In this paper, we take one step towards the distinction between intrinsic and extrinsic motivation. We consider the specific case of Stack Overflow and show that career concerns provide a strong incentive for users to contribute, namely to answer questions posted on the various SO boards. Our strategy for identifying career-concern-based incentives is to estimate the effect of a job change. Our regressions suggest that achieving the goal of a new job leads users to decrease their contribution to SO; and that a drop of about 15% can be assigned to a drop in career concerns. This value is statistically significant and economically significant as well.

Regarding our estimate of the size of the job changing effect, some words of caution are in order. First, our sample results from selection according to a series of criteria. For example, it is likely that the users who choose to link their SO record to their resume are more concerned about their careers than those who keep their SO record unlinked. In this sense, our estimate of career concerns may over-estimate the population average effect. However, the simple theoretical model that forms the basis of our empirical estimation assumes that there are only two states, and that $s = 1$ is an absorbing state. This implies that at $s = 1$ agents have no career concerns at all, which is obviously not very realistic. This in turn suggests that our estimate of career concerns may under-estimate the real value.

An alternative strategy for estimating the career-concerns theory of free user contributions is to directly estimate the probability of a job switch as a function of reputation. (We are currently working on this.) Statistically, one problem with this approach is that easily it suffers from unobserved heterogeneity problems: shocks that change a user’s level of contribution and also lead the user to change jobs may suggest a causality link that does not exist. In principle, a similar objection might be raised regarding our testing strategy. However, it seems more reasonable to assume the change in the unobservable variable impacts the user’s habits at SO before it impacts his or her employment situation.

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.” Working Papers 08-06, NET Institute.
- 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.
- Lerner, Josh and Jean Tirole. 2001. “The open source movement: Key research questions.” *European Economic Review* 45 (4):819–826.
- . 2002. “Some simple economics of open source.” *The Journal of Industrial Economics* 50 (2):197–234.
- Mamykina, Lena, Bella Manoim, Manas Mittal, George Hripcsak, and Björn Hartmann. 2011. *Design Lessons from the Fastest Q&A Site in the West*. New York, New York, USA: ACM.
- Mehra, Amit, Rajiv Dewan, and Marshall Freimer. 2011. “Firms as Incubators of Open-Source Software.” *Information Systems Research* 22 (1):22–38.
- 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.
- Shriver, Scott K, Harikesh S Nair, and Reto Hofstetter. 2013. “Social Ties and User-Generated Content: Evidence from an Online Social Network.” *Management Science* 59 (6):1425–1443.
- Spiegel, Yossi. 2009. “The Incentive to Participate in Open Source Projects: a Signaling Approach.” Working paper.
- Tang, Q, B Gu, and A B Whinston. 2012. “Content Contribution for Revenue Sharing and Reputation in Social Media: A Dynamic Structural Model.” *Journal of Management Information Systems* 29 (2):41–76.
- Varian, H.R. 2012. “Public goods and private gifts.” .

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.